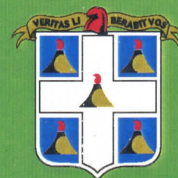
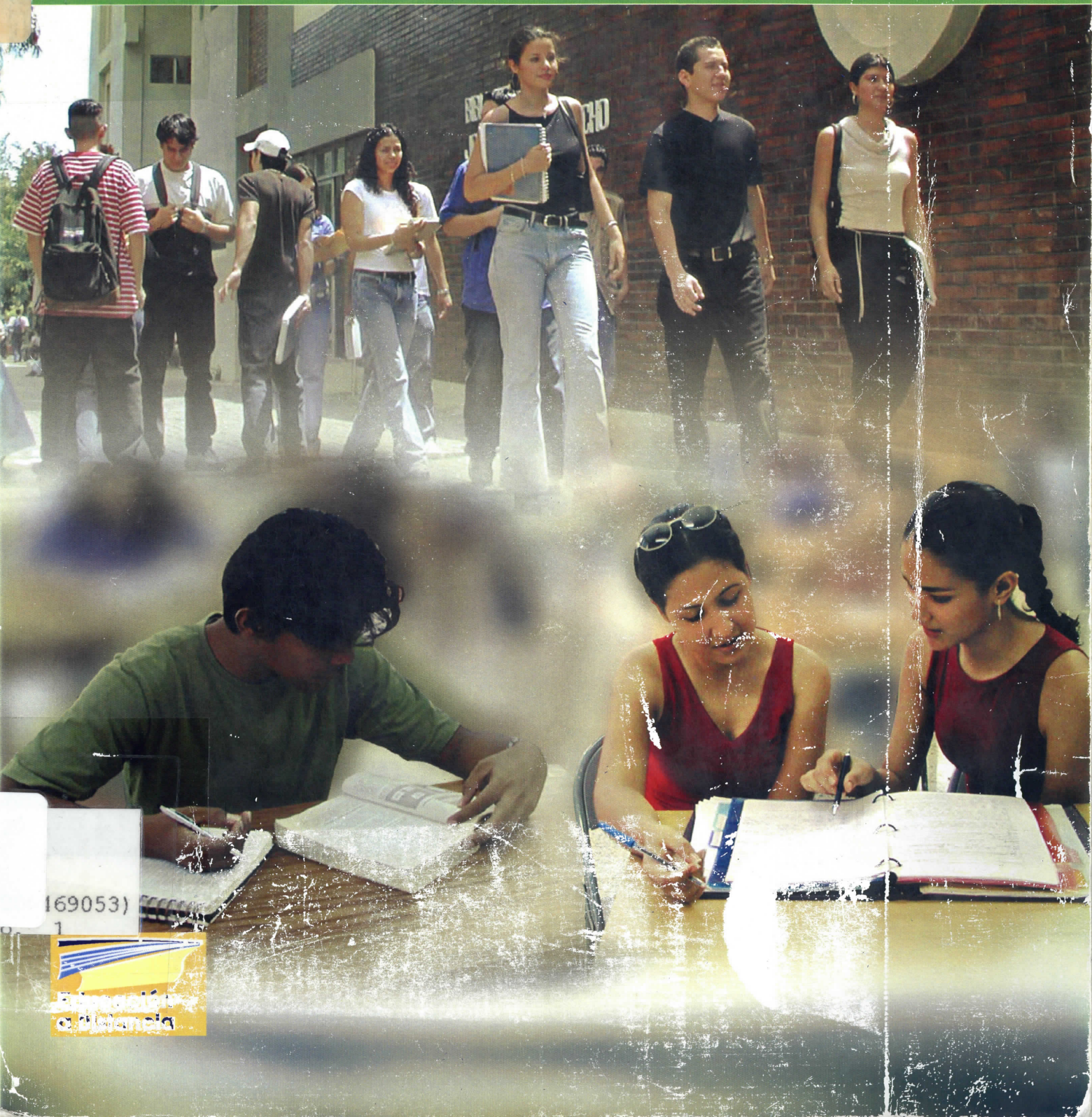


# Técnicas de Recopilación de la Información

## Módulo Autoformativo 11



Universidad  
Centroamericana



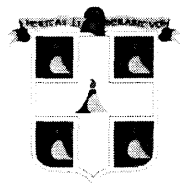
169053)

1





001.4222  
R-934



**UNIVERSIDAD CENTROAMERICANA**  
**Facultad de Ciencias Económicas y Empresariales**  
**Modalidad de Educación a Distancia**

**TÉCNICAS DE RECOPILACIÓN DE LA INFORMACIÓN**  
**MÓDULO AUTOFORMATIVO NO. 11**



## ÍNDICE

<b>Presentación General del Módulo Autoformativo No.11 .....</b>	<b>9</b>
Objetivos generales del módulo .....	12
Esquema de Contenidos del módulo autoformativo 11 .....	13
Orientaciones para el aprendizaje .....	14
Evaluación Diagnóstica .....	15
<b>Unidad Autoformativa I "Generalidades para la recopilación de datos" .....</b>	<b>19</b>
Presentación .....	21
Objetivos de la unidad autoformativa I .....	21
<b>A. LOS DATOS Y EL PROCESO DE MEDICIÓN, MEDICIÓN DE ACTITUDES .....</b>	<b>23</b>
1. La realidad como fuente del conocimiento .....	23
2. La necesidad de datos .....	26
3. El proceso de medición y las escalas de medición .....	28
4. Medición de actitudes y escalas de medición de actitudes .....	32
a. Componente cognoscitivo .....	34
b. Componente afectivo .....	34
c. Componente de comportamiento .....	34
5. La medición de actitudes: Escalas especiales .....	35
a. Escala gráfica .....	36
b. Escala de categorías .....	36
c. Escala de orden jerárquico .....	37
d. Escala de comparación por pares .....	38
e. Escala de suma constante .....	38
f. Escala semántica diferencial .....	39
g. Escala de Stapel .....	39
h. Escala de Likert .....	40
Actividad de Autoaprendizaje No. 1: .....	41
<b>B. POBLACIÓN Y MUESTREO. TIPOS DE MUESTREO: .....</b>	<b>43</b>
1. Concepto de población y muestreo .....	43
a. Concepto de Población: .....	43
b. Concepto de muestra .....	44
c. Leyes del método de muestreo: .....	45
d. ¿Qué es el Muestreo? .....	46
1) Ventajas del Muestreo .....	46
2) Limitaciones del Muestreo .....	47
e. Muestreo y Estadística .....	47
f. Muestreo y Probabilidad .....	48
g. Error de Muestreo .....	48
h. Exactitud y Precisión .....	48
2. Tipos de Muestras o Estrategias de muestreo .....	48
a. Muestras No Probabilísticas .....	49
b. Muestras Probabilísticas .....	50
Actividad de Autoaprendizaje No. 2 .....	52
<b>C. MÉTODOS DE RECOPIACIÓN DE DATOS .....</b>	<b>53</b>
1. Método de observación .....	53
a. Ventajas del método de observación .....	54
b. Desventajas del método de observación .....	54
c. Clasificación de las técnicas de observación .....	54
2. Método de comunicación .....	56



a. Ventajas del método de comunicación.....	56
b. Desventajas del método de comunicación.....	56
c. Clasificación de las técnicas de comunicación.....	57
1) Estructurado - directo.....	58
2) No estructurado - directo.....	59
3) No estructurado-indirecto.....	62
4) Estructurado-indirecto.....	62
3. Medios para recolectar datos por el método de comunicación.....	63
a. Entrevista personal.....	64
b. Entrevista telefónica.....	64
c. Entrevista por correo.....	64
4. Criterios para seleccionar los medios de comunicación.....	65
Actividad de Autoaprendizaje No. 3.....	69
<b>D. ENCUESTAS POR MUESTREO.....</b>	<b>71</b>
1. Planeación de una encuesta por muestreo.....	71
a. Establecimiento de objetivos.....	71
b. Población objetivo.....	72
c. El marco.....	73
d. Diseño de muestreo.....	73
e. Método de medición.....	74
f. Instrumento de medición.....	74
g. Selección y adiestramientos de investigadores de campo.....	75
h. Prueba piloto.....	75
i. Organización del trabajo de campo.....	75
j. Organización del manejo de datos.....	76
k. Análisis de los datos.....	76
2. Diseño de formas (cuestionarios) para la recolección de datos de los encuestados.....	77
a. Los componentes del cuestionario.....	77
b. Diseño del cuestionario.....	79
c. Revisión de las consideraciones preliminares.....	80
d. Decidir sobre el contenido de las preguntas.....	82
e. Decidir sobre el formato de respuestas.....	84
1) Preguntas de respuesta abierta.....	85
a) Ventajas de las preguntas de respuesta abierta.....	85
b) Desventajas de las preguntas de respuesta abierta.....	85
2) Preguntas de selección múltiple.....	86
a) Ventajas de las preguntas de selección múltiple.....	86
b) Desventajas de las preguntas de selección múltiple.....	87
3) Preguntas dicotómicas.....	88
a) Ventajas de las preguntas dicotómicas.....	88
b) Desventajas de las preguntas dicotómicas.....	88
f. Decisión sobre la redacción (las palabras) de las preguntas.....	88
g. Decidir sobre la secuencia de las preguntas.....	92
h. Decidir sobre las características físicas del cuestionario.....	93
i. Llevar a cabo una preprueba, revisión y borrador final.....	93
Actividad de Autoaprendizaje No. 4.....	94
<b>E. ORIENTACIONES PARA LA ELABORACIÓN DEL TRABAJO FINAL.....</b>	<b>95</b>
1. Generalidades.....	95
a. Primera etapa: Elaboración de la propuesta del Plan del Trabajo Final.....	95
b. Segunda etapa: Entrega de la propuesta y revisión por el tutor.....	95



c. Tercera etapa: Ajuste y aplicación del plan, entrega y breve presentación del trabajo final.....	95
2. Sugerencias para la preparación e implementación del plan:.....	96
a. Población Objetivo.....	97
b. Matriz del Estudio.....	98
1) Objetivo General.....	99
2) Objetivos Específicos.....	99
3) Variable.....	99
4) Tipo de Variable.....	100
5) Tipo de escala de medición.....	100
6) Estadígrafo a aplicar.....	101
7) Pregunta.....	101
8) Formato de respuesta.....	101
c. Diseño de muestreo.....	103
1) Marco Muestral.....	103
2) Diseño de muestreo.....	103
3) Método de medición.....	104
d. Prueba Piloto.....	104
e. Preparación del Plan.....	104
f. Implementación del Plan.....	105
g. Documento final.....	105
Actividad de autoaprendizaje 5.....	106
Resumen final de la Unidad Autoformativa I.....	107
Evaluación Final de la Unidad Autoformativa I.....	108
Hoja de respuestas.....	110
Glosario de términos.....	119
Bibliografía.....	122
<b>Unidad Autoformativa II "Muestreo probabilístico aleatorio simple y aleatorio sistemático".....</b>	<b>123</b>
Presentación.....	125
Objetivos de la unidad autoformativa II.....	125
Contenido de la unidad autoformativa II.....	126
<b>A. DISEÑO DE MUESTRAS.....</b>	<b>127</b>
1. Definiciones.....	128
2. Error de estimación en el muestreo.....	131
a. La variación aleatoria.....	131
b. La especificación deficiente de la población.....	132
c. La no-respuesta.....	132
3. Estimaciones en el muestreo.....	134
a. Estimación puntual.....	139
1) Insensamiento.....	140
2) Eficiencia.....	140
3) Consistencia.....	141
b. Estimación por Intervalo.....	141
c. Solución del tamaño de la Muestra:.....	145
Actividad de Autoaprendizaje No. 1.....	148
<b>B. MUESTREO ALEATORIO SIMPLE.....</b>	<b>149</b>
1. Definición.....	149
2. Procedimiento de selección de una muestra aleatoria simple.....	150
3. Estimaciones basada en una muestra aleatoria simple.....	153
4. Estimador de la media poblacional $\mu$ , varianza estimada de la media y	



límite para el error de estimación de la media .....	154
5. Estimador del total poblacional $\tau$ , varianza estimada del total y límite para el error de estimación del total poblacional.....	158
6. Estimador de la proporción poblacional p, varianza estimada de la proporción y límite para el error de estimación de la proporción poblacional .....	160
7. Cálculo del tamaño de la muestra para la media poblacional, el total poblacional y la proporción poblacional .....	161
a. Tamaño de la muestra para la media poblacional ( $\mu$ ) .....	161
b. Tamaño de la muestra para estimar el total poblacional ( $\tau$ ) .....	162
c. El tamaño de la muestra para estimar la proporción poblacional ( $p$ ) .....	163
Actividad de autoaprendizaje No. 2.....	164
<b>C. MUESTREO ALEATORIO SISTEMÁTICO .....</b>	<b>167</b>
1. Definición .....	167
2. Procedimiento de selección de una muestra aleatoria sistemática .....	167
a. Ventajas de la selección sistemática.....	168
b. Desventajas de la selección sistemática .....	169
3. Definición de poblaciones aleatorias, ordenada, periódica y su efecto en la varianza y las estimaciones .....	169
4. Estimaciones basadas en una muestra aleatoria sistemática para poblaciones aleatorias.....	171
5. Estimador de la media poblacional $\mu$ , varianza estimada de la media y límite para el error de estimación de la media .....	171
6. Estimador del total poblacional $\tau$ , varianza estimada del total y límite para el error de estimación del total poblacional.....	173
7. El estimador de la proporción poblacional p, varianza estimada de la proporción y límite para el error de estimación de la proporción poblacional. ....	174
8. Cálculo del tamaño de la muestra para la media poblacional, el total poblacional y la proporción poblacional .....	175
a. Tamaño de la muestra para estimar la media poblacional ( $\mu$ ) .....	175
b. Tamaño de la muestra para estimar el total poblacional ( $\tau$ ) .....	176
c. Tamaño de la muestra para estimar la proporción poblacional ( $p$ ) .....	176
Actividad de autoaprendizaje No. 3.....	178
Resumen final de la unidad autoformativa II .....	179
Evaluación Final de la unidad autoformativa II .....	180
Hoja de respuestas .....	182
Glosario de términos .....	184
Bibliografía.....	186
Apéndice Unidad Autoformativa II .....	187
<b>Unidad Autoformativa III “Muestreo probabilístico aleatorio estratificado y aleatorio por conglomerado”.....</b>	<b>189</b>
Presentación.....	191
Objetivos de la unidad autoformativa III .....	191
Contenido de la unidad autoformativa III.....	192
<b>A. MUESTREO ALEATORIO ESTRATIFICADO .....</b>	<b>193</b>
1. Definición .....	193
a. Ventajas del muestreo aleatorio estratificado.....	193
b. Criterios de estratificación.....	194

2. Procedimiento de selección de una muestra aleatoria estratificada .....	195
3. Estimaciones basadas en una muestra aleatoria estratificada .....	197
4. Estimador del total poblacional $\tau$ , varianza estimada del total y límite para el error de estimación del total poblacional .....	199
5. Estimador de la proporción poblacional $p$ , varianza estimada de la proporción y límite para el error de estimación de la proporción poblacional .....	200
6. Cálculo del tamaño de la muestra para la media poblacional, el total poblacional y la proporción poblacional .....	202
Actividad de autoaprendizaje No. 1 .....	207
<b>B. MUESTREO ALEATORIO POR CONGLOMERADOS .....</b>	<b>209</b>
1. Definición .....	209
2. Procedimiento de selección de una muestra aleatoria por conglomerado .....	209
3. Estimaciones basada en una muestra aleatoria por conglomerado .....	211
4. Estimador de la media poblacional $\mu$ , varianza estimada de la media y límite para el error de estimación de la media .....	212
5. Estimador del total poblacional $\tau$ , varianza estimada del total y límite para el error de estimación del total poblacional .....	214
6. Cálculo del tamaño de la muestra para la media poblacional y el total poblacional .....	216
7. Estimador de una proporción poblacional, varianza estimada de la proporción y límite para el error de estimación de la proporción poblacional .....	218
8. Cálculo del tamaño de la muestra para las proporciones: .....	220
Actividad de autoaprendizaje No.2 .....	221
Resumen final de la Unidad Autoformativa III .....	222
Evaluación Final de la Unidad autoformativa III .....	223
Hoja de respuestas .....	225
Glosario de términos .....	226
Bibliografía .....	227

6847

Ingreso	28-01-04
Comprado a	
Donado por	Edic. a Dist.
Precio	Reg 200464053





***Presentación General***  
***del Módulo Autoformativo No.11***



**Bienvenido futuro profesional:**

*El presente módulo autoformativo de Técnicas de Recopilación de la Información le permitirá introducirse en el diseño de muestras probabilísticas que le serán de utilidad en el proceso de investigación. Retoma, complementa y profundiza los conocimientos adquiridos en la asignatura de Estadística, a partir de la información que se recolecta y procesa en las muestras; con el propósito de realizar inferencias sobre las poblaciones que son objeto de estudio y determinar los tamaños de muestras requeridos por el método de investigación, constituyéndose, entonces, en uno de los pasos que exige la aplicación del método científico.*

*El aprendizaje de los contenidos de este módulo contribuirá a que logre el cumplimiento del rasgo del perfil de la carrera de Administración de Empresas, relacionado con la actividad investigativa, la que irá perfeccionando a lo largo de su formación profesional, con el estudio de otros módulos.*

*Técnicas de Recopilación de la Información tiene como requisito la asignatura de "Estadística" ya estudiada en el cuatrimestre anterior y a su vez es prerequisite de "Técnicas de Investigación" que se estudiará posteriormente.*

*También los aprendizajes logrados, en esta asignatura le servirán de base para el estudio del módulo de "Investigación de Mercados" en el que aplicará todo lo aprendido al realizar investigaciones de mercado, que son de mucho interés para sustentar con acierto, la toma de decisiones en la empresa.*

*Los conocimientos, habilidades, destrezas y valores éticos que desarrolle en el presente módulo le serán útiles para realizar investigaciones en asignaturas de estudio paralelo o posterior, ya que la práctica investigativa es una constante en su proceso de formación como profesional de la administración, y en el ejercicio profesional, una vez graduado.*

*Para aprovechar al máximo el presente módulo, es necesario que domine tanto el significado como las operaciones de cálculo, entre ellas, medidas de tendencia central, medidas de dispersión, probabilidades, distribuciones de probabilidad y estimaciones puntuales y por intervalo. También es indispensable que le dedique tiempo suficiente necesario al auto estudio y desarrolle las actividades de autoaprendizaje y evaluación para así poder culminar con éxito el esfuerzo que usted y la UCA nos hemos propuesto.*



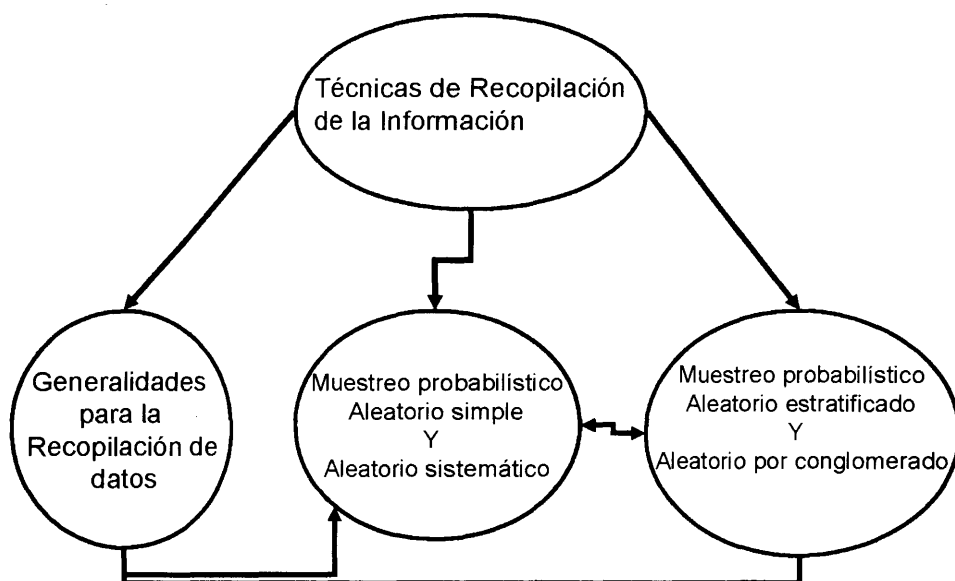
## **Objetivos generales del módulo**

1. Identificar los métodos de recopilación de la información para la realización de investigaciones propias en la administración y economía.
2. Aplicar adecuadamente los elementos técnicos que se requieren para el diseño de encuestas, orientadas a la recopilación de información relevante.
3. Diferenciar los diversos métodos para la selección de una muestra probabilística, identificando las situaciones en que es adecuada su aplicación.
4. Calcular con exactitud las estimaciones de la media poblacional, total poblacional, proporción poblacional así como el tamaño de la muestra para los distintos diseños de muestreo probabilístico estudiados.
5. Desarrollar las habilidades de un pensamiento crítico y de análisis de datos de una realidad particular por medio de la elaboración de un trabajo final en donde aplique todo lo aprendido en el presente módulo.



## Esquema de Contenidos del módulo autoformativo 11

Para que usted se forme una visión general de los contenidos del módulo de técnicas de recopilación de la información, le presentamos una descripción esquemática de las tres unidades autoformativas que lo constituyen y de las interrelaciones que guardan.



### **Primera unidad autoformativa: Generalidades para la recopilación de datos.**

Esta unidad nos permitirá conocer los distintos métodos para recopilar datos en la investigación, concentrando el estudio en las encuestas y el diseño de instrumentos o cuestionarios como método de gran utilidad en las ciencias administrativas y económicas.

En esta unidad complementaremos lo estudiado en el módulo de Estadística relacionado con las escalas de medición, concentrando el estudio en la medición de actitudes.

Así mismo revisaremos y profundizaremos el concepto de muestreo y los tipos de muestreo para dar paso al aprendizaje de las siguientes dos unidades autoformativas en las que estudiaremos con profundidad los cuatro tipos de diseños de muestras probabilísticas.

### **Segunda unidad autoformativa: Muestreo probabilístico aleatorio simple y aleatorio sistemático.**

En esta segunda unidad estudiaremos los primeros dos métodos de muestreo aleatorio, el simple y el sistemático en el que aplicaremos diversas fórmulas para realizar el cálculo de estimaciones de la media poblacional, total poblacional, proporción poblacional y tamaño de la muestra de los métodos arriba señalados.

Esta temática complementará aspectos que hemos estudiado en estadística relacionados con la estimación puntual y por intervalo.

**Tercera unidad autoformativa: Muestreo probabilístico aleatorio estratificado y aleatorio por conglomerados.**

En ella estudiaremos los otros dos métodos de muestreo aleatorio, el estratificado y el de conglomerado o de racimo. Aplicaremos fórmulas para realizar el cálculo de estimaciones de la media poblacional, total poblacional, proporción poblacional y tamaño de la muestra de estos otros dos métodos.

## **Orientaciones para el aprendizaje**

El módulo de Técnicas de Recopilación de la Información, está estructurado de forma tal que contribuye a su proceso de autoaprendizaje, característica esencial de la modalidad educativa a distancia que desarrolla la Universidad Centroamericana.

Durante el estudio de sus tres unidades autoformativas, usted irá encontrando detalladamente cada uno de los elementos que la integran. Estos elementos son:

- La presentación general del módulo, que le permite visualizar un panorama general de la temática por abordar.
- Los objetivos generales del módulo y de cada unidad autoformativa en particular lo que representa la orientación de los aprendizajes que deseamos alcanzar una vez se concluya cada unidad autoformativa y
- El módulo en general.

Cada tema se inicia con la presentación y explicación de los aspectos teóricos, con ejercitaciones que refuerzan lo planteado y vinculan lo teórico con lo práctico, esto es de vital importancia para el proceso de aprendizaje de conocimientos y el desarrollo de habilidades y actitudes relacionadas con la temática estudiada.

Como parte indispensable para su autoevaluación continua, se presentan actividades de autoaprendizaje, diseñadas en correspondencia con el tema de estudio. En ellas aplicará lo aprendido y lo vinculará con las situaciones concretas del entorno en donde desarrolla sus actividades cotidianas.

Con la estructura así definida, podrá organizar su estudio, sin tener que preocuparse por la coherencia de la enseñanza dado que cada unidad ésta diseñada de tal manera que usted puede llevar una secuencia lógica y pedagógica del desarrollo de su propio aprendizaje.

Las actividades de autoevaluación continua, le servirán para verificar sus logros y la detección de sus errores, cuando responda a los cuestionarios, resuelva los problemas planteados y compare sus resultados con los presentados en la hoja de respuesta correspondiente.

Favor escriba sus dudas y las dificultades que logre encontrar en el proceso de estudio y pregúntelas a su tutor(a) en el período de consulta establecido en el encuentro tutorial. El (ella) le ayudará a aclarar sus dudas y superar los inconvenientes que se presenten.

Si usted sigue paso a paso las orientaciones del módulo y las que su tutor(a) complemente en los periodos asignados a consultas, debates, demostraciones, investigaciones y otras actividades grupales, es seguro que el aprendizaje autónomo será para usted una agradable experiencia y habrá conseguido un excelente rendimiento y dominio del contenido que le será de mucha utilidad para su formación profesional.

## **Evaluación Diagnóstica**

Busque las respuestas a esta evaluación diagnóstica en las páginas 110, 111, 112 y 113, al final de la unidad autoformativa I, corrija y valore cuánto énfasis y tiempo debe dedicarle al estudio y revisión de los conocimientos previos necesarios.

1. Realice un esquema donde se reflejen los tipos de variables estudiadas en estadística y brinde dos ejemplos de cada una de ellas.
2. Indique la tipología y el dominio de las siguientes variables: (Nota: El dominio es el conjunto de valores que puede tomar la variable, también se le conoce con el nombre de soporte).
  - a. Número de huevos puestos por las gallinas cada día en una granja.
  - b. Temperatura registrada cada media hora en un observatorio.
  - c. Medida – grande, mediana o pequeña – de las hojas de una determinada especie de planta.
  - d. Número de habitantes de los municipios de Managua.
  - e. Número de hijos de una familia en el Municipio del Tuma – La Dalia.
  - f. Edad de los alumnos de Técnicas de Recopilación del grupo de Educación a distancia al que pertenezco.
  - g. Nivel de infección de una planta por un determinado tipo de hongo.
  - h. Número de cajillas de gaseosas producida diariamente por Kola Shaler.
  - i. Percepción del servicio de transporte público en Managua.
  - j. Color de los ojos de los/las estudiantes del grupo de educación a distancia al cual pertenezco.
3. Realice un esquema donde se reflejen los tipos de escalas o niveles de medición estudiados en estadísticas y brinde un ejemplo de cada una de ellas.
4. Anote el nivel de medición de las variables que se presentan en el punto número 2.
5. Se pide lo siguiente:
  - a) calcule la media de los siguientes valores muestrales: 5, 9, 4, 10.
  - b) Demuestre con los datos obtenidos en a), que  $\sum (x - \bar{x}) = 0$ .



6. Una empresa de servicios de energía eléctrica seleccionó 20 clientes residenciales. Los siguientes son los importes en córdobas que se cargó a los clientes por el servicio eléctrico el último mes:

54	48	58	50	25	47	75	46	60	70
67	68	39	35	56	66	33	62	65	67

- Calcule el importe mensual promedio
  - ¿Es éste un dato estadístico muestral o un parámetro poblacional?
7. Las edades de una muestra de pasajeros que viajan en el Expreso de Matagalpa a Managua fueron: 32, 21, 60, 47, 54, 17, 72, 55, 53, y 41.
- Calcule la amplitud total.
  - Calcule la desviación media
  - Calcule la desviación estándar
  - Anote el valor de la media?
  - Anote el valor de la varianza?
8. Calcule la varianza y la desviación estándar de los siguientes pesos muestrales: 7, 9, 11, 9, y 4 gramos.
9. Antes de efectuar una encuesta a nivel nacional, se seleccionaron 40 personas para probar el cuestionario. Una pregunta acerca de si debe o no legalizarse el aborto, requiere una respuesta de sí o no.
- ¿Cuál es el experimento?
  - Enuncie un posible evento
  - Diez de las 40 personas se declararon a favor de legalizar el aborto. Con base en estas respuestas muestrales, ¿cuál es la probabilidad de que una persona específica esté a favor de tal legalización?
  - ¿Qué concepto de probabilidad ilustra esto?
  - ¿Los eventos son por igual probables, mutuamente excluyentes y colectivamente exhaustivos?
10. Calcule la media y la varianza de la siguiente distribución probabilística discreta:

x	P(x)
0	0.20
1	0.40
2	0.30
3	0.10

11. Una industria produce cojinetes de bolos en forma automática en una máquina Tronar BBX. Para uno de los cojinetes, la media aritmética de los diámetros se determina como 20.00mm. La desviación estándar de la producción durante un largo periodo se calcula como 0.150mm.
- ¿Qué porcentaje de los cojinetes tendrán diámetros entre 20.00 y 20.27 mm.?
  - ¿Qué porcentaje de los cojinetes tendrán diámetros de 20.27mm. o más?
  - ¿Qué porcentaje de los cojinetes tendrán diámetros entre 19.85mm. y 20.30 mm.?
  - ¿Qué porcentaje de los cojinetes tendrán diámetros de 19.91mm. o menos?

Nota: Este ejercicio se resuelve utilizando la distribución probabilística normal estándar.

12. Una muestra de 10 observaciones se selecciona a partir de una población normal para la cual se conoce que la desviación estándar es 5. La media muestral tiene un valor de 20.
- Determine el error estándar de la media
  - Explique porqué se puede utilizar la fórmula:  $\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$ , para determinar el intervalo de confianza de 95% aunque el tamaño de la muestra es menor que 30.
  - Determine el intervalo de confianza de 95% para la media de la población. Nota: Este ejercicio corresponde a la estimación por intervalo con un nivel de confianza del 95%.

***Unidad Autoformativa I***  
***“Generalidades para la  
recopilación de datos”***



## **Presentación**

*La presente unidad autoformativa le permitirá identificar que para la realización de investigaciones en las ciencias sociales o ciencias de la conducta a la cual pertenece la administración y la economía es de suma utilidad la técnica de las encuestas por muestreo.*

*En esta unidad, también podrá diferenciar los tipos de muestras que se pueden aplicar, así como el diseño de los cuestionarios que son los instrumentos que permiten la recolección de la información para su posterior procesamiento y análisis, debiendo tener mucho cuidado en las escalas de medición de las variables que estamos interesados en estudiar en la población objeto de estudio.*

*Al final de la unidad encontrará las orientaciones para que elabore, con los conocimientos adquiridos, un plan para el estudio por encuestas, de una situación determinada que se completará con las dos unidades autoformativas restantes del presente módulo.*

## **Objetivos de la unidad autoformativa I**

1. Comprendo la importancia de los datos, el proceso de medición y la medición de actitudes en la investigación.
2. Establezco las diferencias entre las muestras probabilísticas y no probabilísticas en los estudios de investigación.
3. Diferencio los métodos de recopilación de datos en las investigaciones.
4. Identifico diversas pautas que me permitan realizar el proceso de planeación de una encuesta y el diseño del instrumento.
5. Elaboro un plan para la realización de un estudio de encuesta por muestreo que presentaré como trabajo final del módulo.

## A. LOS DATOS Y EL PROCESO DE MEDICIÓN, MEDICIÓN DE ACTITUDES

---

### 1. La realidad como fuente del conocimiento

Recuerdo que en secundaria estudié la ley de la caída libre de los cuerpos. Esta ley me dice que todos los cuerpos caen al suelo con la misma velocidad, bajo las mismas condiciones externas, que no afecten la caída libre.

Puedo realizar un experimento para comprobar esta ley, ubicándome en un lugar donde no tenga la influencia de aire y puedo verificar que un gramo de papel, cae con la misma velocidad que un gramo de acero, al ser dejado libre a una misma altura.

Recuerdo las ecuaciones que estudié, la velocidad en caída libre, que está dada por  $v = g * t$ , donde  $v$ : es la velocidad con que cae el objeto,  $t$ : es el tiempo de caída y  $g$ : es la aceleración de la gravedad, que es el nivel de aceleración que le imprime la tierra a todos los cuerpos que se encuentran en su campo de gravedad y que es considerada una constante,  $g = 9.81 \text{ m/seg.}^2$

Observo que al estar dada la velocidad por  $v = g * t$ , la velocidad es independiente de la masa del cuerpo, es así que entonces al dejar caer un gramo o una tonelada de papel o de acero desde una misma altura, bajo condiciones iguales y sin influencia de otras fuerzas o elementos que incidan en los cuerpos, estos caerán con la misma velocidad.

Para comprender lo anterior, es conveniente que recordemos la expresión  $F = m * a$ , donde  $F$ : es la fuerza que se le aplica a un determinado cuerpo,  $m$ : la masa del cuerpo y  $a$ : la aceleración que se le imprime al cuerpo provocada por la fuerza, dada una masa determinada. Como podemos observar  $a$  es el equivalente a  $g$ , la aceleración que le imprime la fuerza de atracción de la tierra al cuerpo.

Cuando el cuerpo posee una masa mayor, la fuerza que se le aplica por la atracción terrestre es mayor si la comparamos con la fuerza que le aplica la tierra a un cuerpo de masa menor. En otras palabras, cuando la masa (masa inercial) es mayor, la fuerza será mayor, siendo la aceleración de la gravedad constante para todos los cuerpos que caen del reposo a una determinada altura.

La ley anterior ha sido formulada en el siglo XVI por el físico Isaac Newton y forma parte de un grupo de leyes que se conocen como ley de la gravitación universal y que también se relacionan con las famosas leyes de Newton para la Mecánica: Ley de la inercia de los cuerpos, ley de la fuerza y ley de la acción y reacción.

Antes de la formulación de sus leyes, Newton solía sentarse a descansar y apreciar la naturaleza debajo de un árbol de manzana, ahí observó muchas veces la caída de manzanas de forma natural, esto le llevó a un proceso de reflexión y de cuestionamiento del ¿Porqué? Caían las manzanas.

De forma similar, Charles Darwin por medio de la observación de la Naturaleza, de plantas y animales formuló la teoría de la evolución de las especies.

En la actualidad por ejemplo se desarrollan proyectos de investigación que analizan el ADN humano como el proyecto del Genoma Humano, que se orienta a descifrar la información genética de nuestros núcleos celulares. Con el conocimiento de la secuencia completa del genoma humano aprenderemos más acerca de nosotros mismos y también de las enfermedades hereditarias.

Desde todas las disciplinas científicas, la humanidad busca la ampliación del conocimiento de la Naturaleza y de nosotros mismos.

### ***Pero ¿Cómo se realiza ese conocimiento?***

De forma muy breve, y apoyándonos en la figura 1, por un lado me muestra a la Realidad concreta, la Naturaleza, y todo lo que me rodea; por el otro lado me encuentro Yo, como especie humana y que puedo considerarme como parte de la Realidad, de la Naturaleza (observador).

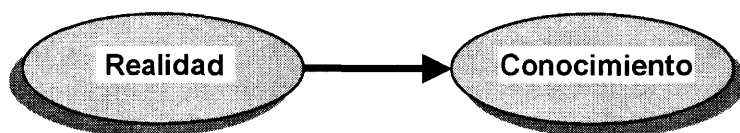


Figura 1

La naturaleza y todo lo que me rodea, de forma permanente se me está mostrando por medio de diversas señales. Esas señales son percibidas por mi parte, por medio de los sentidos sensoriales de la vista, el oído, el gusto, el tacto y el sentir.

Estas sensaciones percibidas originan en el cerebro un proceso de abstracción de esa realidad, haciéndome preguntas, tratando de entender y desentrañar el mundo que me rodea, por penetrar en sus relaciones y descubrir el posible sentido de las cosas que existen a mi alrededor, buscando respuestas a mis interrogantes, todo esto me conduce a lo que se le denomina como conocimiento.

El conocimiento entonces lo puedo definir como el conjunto de abstracciones de la realidad que me aseguran comprenderla y poder incluso con mi acción modificarla.

El conocimiento lo puedo tipificar de varias formas, dos de ellas que me permiten comparar es el conocimiento vulgar y el científico. El conocimiento vulgar, es una forma de conocimiento práctico que se transmite directamente de unos a otros y se manifiesta, en parte, en la cultura popular, es un pensar espontáneo presente en la vida cotidiana.

Por el contrario, el conocimiento científico es más afinado que el vulgar porque permite aventurarnos a las conjeturas y criticar las explicaciones logradas a través del sentido común, por su extensión, claridad, precisión, por centrarse más en el porqué de las cosas y por su carácter sistemático y seguridad comprobada.

El conocimiento científico queda caracterizado por el método que utiliza y por ello puede definirse como todo conocimiento adquirido por medio del método científico, este tipo de conocimiento es el que está más vinculado a la investigación.



### ***Pero ¿Cómo la investigación científica acerca la realidad con el conocimiento?***

Para ello establezco algunos conceptos básicos que maneja la ciencia, apoyándome en la figura 1 que me relaciona la realidad con el conocimiento, expandiéndola y obteniendo la figura 2.



Figura 2

En la figura 2, se muestra que la realidad es la fuente de los hechos, estos se nos muestran en fenómenos y de los fenómenos recolectamos datos, información.

Un hecho es todo aquello que se sabe o se supone, con algún fundamento, que pertenece a la realidad. Por ejemplo: la vocalización de una palabra, un aula, un profesor explicando una clase. Como puedo apreciar, existe una gran variedad de hechos, que a su vez pueden considerarse como acontecimientos, procesos y sistemas concretos que pueden ser de interés para su estudio.

Los hechos que tienen lugar en el espacio y en el tiempo, como pueden ser un grito, un golpe, una pregunta o una respuesta, se denominan acontecimientos. Cuando varios acontecimientos constituyen una secuencia temporalmente ordenada, de tal manera que cada acontecimiento implica, afecta o permite los siguientes, estamos ante un proceso. Así, en un proceso didáctico podrían estar implicados acontecimientos como la explicación del profesor y las anotaciones y respuestas del estudiante. Ahora bien, en sentido estricto, podemos considerar que la mayoría de los acontecimientos están constituidos, a su vez, por procesos. Por ejemplo, un profesor explicando puede considerarse como un acontecimiento y a su vez como un proceso de acontecimientos si consideramos que pronuncia palabras, las escribe en la pizarra y son percibidas por los estudiantes.

Por último, cuando un hecho está constituido por partes identificables y estructuradas, formando un todo unitario, estoy ante la presencia de un sistema concreto, como puede ser un aula de clase, un taller de pintura, un equipo audiovisual o un claustro de profesores.

Observemos que en un aula de clase las partes identificables y estructuradas se asocian a personas (estudiantes y docentes), medios de enseñanza- aprendizaje como pueden ser el material autoformativo en nuestro caso, retroproyector, filminas, entre otros; así como otros elementos o partes identificables que se asocian al objetivo de la clase y el plan de clase entre otros, así como también las relaciones que se presentan entre las partes.

Para el taller de pintura y el equipo audiovisual ¿Qué partes identificables y estructuradas se asocian? Me preparo para compartir con mis compañeros las partes identificables y estructuradas para el taller de pintura y el equipo audiovisual.

Para conocer y dominar la realidad que le rodea, el ser humano percibe y capta, por medio de sus sentidos, una gran variedad de hechos que constituyen acontecimientos, procesos y sistemas concretos.

Cuando son percibidos por el investigador/observador, los hechos observables suelen denominarse fenómenos: es decir, son captados por el observador y percibidos a través de los sentidos como fenómenos; en consecuencia, el fenómeno designa un hecho percibido.

El conocimiento que se extrae de los fenómenos lo denominamos información; ésta se transmite en forma de datos, que es una información fijada o codificada por el investigador.

Cualquier información, por pequeña o fragmentaria que sea, puede considerarse como un dato. Por tanto, el dato encierra enunciados, afirmaciones o negaciones que expresan aspectos o características concretas de la realidad (Sierra Bravo, 1984, 121). Por ejemplo, el sujeto X es niño, tiene 10 años y obtuvo una puntuación de 100 en un test de inteligencia.

A manera de síntesis, el conocimiento es generado por la propia realidad que impacta en nosotros los seres humanos por medio de los sentidos, generándose un proceso de abstracción de esa realidad que nos hace inquietarnos para poder conocer y explicarnos los hechos que acontecen en la realidad.

De lo anterior aprecio la importancia que en la investigación tienen tanto los datos como la necesidad de ellos y las formas como los puedo recolectar, estos últimos temas son los que estaremos estudiando en la presente unidad autoformativa.

***Me ejercito:***

Elaboro un ejemplo de mi propia experiencia en el que se cumplan los componentes de la figura 2. Me preparo para compartir con mis compañeros.

## **2. La necesidad de datos**

A manera de ejemplo, me encuentro interesado en conocer si la empresa en la que laboro esta obteniendo ganancias o pérdidas?. ¿Cómo puedo darme cuenta?. Acaso me doy cuenta solamente observando los movimientos que realiza el personal y por la cantidad de productos que se producen y se venden?

*¡La verdad es que no!*

Para poder darme cuenta, es conveniente que en la empresa se lleve un exhaustivo registro de datos contables que me permita la posibilidad de procesarlos para poder llegar a la conclusión de que estamos o no ganando en la empresa.

Debemos de llevar entonces un registro de todas las transacciones económicas, registrarlas adecuadamente por tipo de cuenta en el sistema de registro contable de la empresa, procesar posteriormente todos estos datos para consolidarlos en los estados financieros (Balance general y balance de pérdida y ganancias) que hemos estudiado en

contabilidad I, lo que nos dará una imagen de la dimensión financiera que posee la empresa en un determinado período.

De igual manera para realizar un estudio o investigación, requiero de datos que sean los que de forma pertinente estén asociados a lo que deseo estudiar, que me permita detectar las características que se nos manifiestan en los fenómenos que estudiamos y que se asocian a los hechos que nos presenta la realidad, por lo que defino un dato como:

Un dato es cada uno de los elementos de información que se recoge durante el desarrollo de una investigación y sobre la base de los cuales, convenientemente sintetizados, podrán extraerse conclusiones en relación con el problema inicial que estoy interesado en estudiar.

Un dato es por ejemplo, saber que la persona X opina que las pruebas nucleares deben ser proscritas.

Esa información, por sí sola, carece prácticamente de valor, pues poco nos dice de las reacciones que despiertan las pruebas de armas atómicas en la gente. Pero el valor del dato reside no en su alcance individual, en lo que nos expresa por sí mismo, sino en su posibilidad de ser integrado en conjuntos mayores.

Cuando agrupamos muchas informaciones de carácter similar, cada dato se hace valioso dentro de una perspectiva más amplia.

Así, en nuestro ejemplo, si consultamos la opinión de muchas personas, podemos llegar a enunciar que un determinado tanto por ciento de ellas están en contra de los ensayos nucleares e integrar esa información, a su vez, en un estudio sobre las opiniones de determinado conglomerado social.

Las fuentes de datos pueden ser personas, situaciones o hechos que se observan directamente, o materiales bibliográficos de diversa naturaleza. Las llamamos unidades de datos y, a su conjunto, a la suma de todas las unidades, se les da el nombre de universo o población.

En general, toda investigación puede considerarse como una búsqueda de los datos apropiados que permitan resolver ciertos problemas de conocimiento. Estos datos son obtenidos a través de un conjunto de unidades que constituyen el universo relevante para la investigación.

Existen universos que resultan demasiado amplios para el investigador, pues éste no tiene ni el tiempo ni los recursos para abordar el estudio de cada una de las unidades que lo componen por ejemplo: el conjunto de ciudadanos de un país, la flora de una región o las innumerables galaxias.

De igual manera, el conjunto total de hechos y fenómenos que la realidad nos presenta, no puede ser obtenida totalmente para realizar una investigación debido a la infinitud tanto del Universo como a la limitante real de los seres humanos.

Para resolver este inconveniente, se acude a la selección de una parte de los elementos del universo o población mediante la extracción de muestras, esta es una parte de la población o del universo que estudiaré en el tema B del presente módulo autoformativo.

La ciencia entonces estudia un fenómeno  $x$  y trata de establecer leyes generales que expliquen el comportamiento de la realidad para la mayor parte de los fenómenos  $x$ 's que se presenten, es así que de la caída de la manzana Isaac Newton logró identificar y formular una ley general que explica la caída libre de cualquier cuerpo afectada por el campo gravitacional de la Tierra.

Lo anterior no debe verse como que estas leyes que se logra formular sean inmutables, al contrario, 'se someten a revisión, contrastación, y mayor generalización de forma permanente y por siempre: esta es la razón de la ciencia.

Continuando con los datos, no se nos presentan por sí solos, sino que deben ser obtenidos de los fenómenos y medidos adecuadamente para poderlos procesar posteriormente.

Pero ¿Cómo podemos medir los fenómenos?. Esta pregunta nos genera el estudio y significado de lo que es medición y el proceso de medición que ya estudiamos en el módulo de Estadística y conviene que lo tengamos presente.

**Me ejercito:**

Identifico de mi propia experiencia algunos ejemplos de datos y los comparto con mis compañeros.

### 3. El proceso de medición y las escalas de medición

El proceso de medición es un aspecto fundamental de la investigación. Con frecuencia se afirma que para entender realmente una cosa lo mejor es tratar de medirla.

Lord Kelvin, famoso físico inglés y que en su honor conocemos la escala de medición de temperatura denominada escala "Kelvin", decía:

*"Si puede usted medir de lo que habla y expresarlo con un número, sabe algo de su tema; pero si no puede medirlo, sus conocimientos son de una especie pobre y muy poco satisfactorios".*

En el módulo de estadística hemos estudiado lo que es el proceso de medición, a manera de síntesis, estudiamos la asociación del sistema empírico con el sistema abstracto y más concretamente la asociación con el sistema numérico, habiendo estudiado las cuatro escalas de medición básicas que son: Ordinal, Nominal, de Intervalo y de Razón

La relación existente entre las escalas de medición, las propiedades del sistema numérico y las estadísticas que son apropiadas para cada tipo de escala son resumidas en el cuadro 1.

**Características de las escalas de medición**  
**Cuadro 1**

Escala	Sistema Numérico	Estadística Permisible
Nominal	Definición única de numerales (0, 1, 2, ..., 9)	Porcentajes Moda Prueba binomial Prueba de Ji-cuadrada
Ordinal	Orden de los numerales ( $0 < 1 < 2 \dots < 9$ )	Percentiles Mediana Correlación de Rango - Orden
Intervalo	Igualdad de diferencias ( $2-1=4-3$ )	Rango Media Desviación estándar Correlación Producto- Momento
Razón	Igualdad de razones ( $2/4=4/8$ )	Media geométrica Media armónica Coeficiente de variación

Observemos que la segunda columna nos sintetiza la característica del sistema numérico que cada escala nos permite y en la tercera columna nos relaciona la estadística que el tipo de escala nos permite que apliquemos. Las estadísticas que le son aplicables a escalas ubicadas en posición de arriba son aplicables a las escalas en posición de abajo, pero lo contrario no es aplicable.

Complementando lo anterior, en el cuadro 2, se muestran las características de las escalas comparadas con las características del sistema numérico

**Características de las escalas de medición versus el sistema numérico**  
**Cuadro 2**

Escala	Característica de la escala	Característica del sistema numérico
Nominal	<ul style="list-style-type: none"> <li>Clasifica los objetos por categorías.</li> <li>Los objetos son mutuamente excluyentes.</li> </ul>	<ul style="list-style-type: none"> <li>Los números sólo reflejan que los objetos difieren entre sí.</li> <li>Las únicas operaciones aritméticas que se pueden utilizar son la igualdad y desigualdad.</li> </ul>
Ordinal	<ul style="list-style-type: none"> <li>Emplea números para ordenar objetos respecto a cierta característica.</li> <li>No permite determinar la distancia entre la medida de los objetos con respecto a la característica que se mide.</li> </ul>	<ul style="list-style-type: none"> <li>Se pueden utilizar las operaciones de igualdad y desigualdad.</li> <li>Las operaciones de menor que (<math>&lt;</math>) y mayor que (<math>&gt;</math>) se pueden utilizar.</li> </ul>
Intervalo	<ul style="list-style-type: none"> <li>Ordena los objetos de manera que las diferencias entre los valores de la escala sean iguales, porque tiene una unidad de medición.</li> </ul>	<ul style="list-style-type: none"> <li>Las operaciones aritméticas que podemos usar son: Igualdad, desigualdad, de ordenación y sustracción.</li> </ul>
Razón	<ul style="list-style-type: none"> <li>Similares a la de intervalo en que poseen una unidad de medición, pero además un significativo punto cero.</li> </ul>	<ul style="list-style-type: none"> <li>Todas las operaciones aritméticas son significativas: Igualdad, desigualdad, ordenación, sustracción, adición y división.</li> </ul>

Debemos tener presente estas características del sistema numérico cuando asociemos una escala a la característica estadística que estudiamos, con el objeto de minimizar los errores de medición. Favor revisar el material autoformativo de estadística para tener presente estos conceptos que serán considerados en la orientación para la preparación del trabajo final.

Existen dos conceptos que son de mucha importancia cuando se estudia la medición: la validez y la confiabilidad, los que debemos tener presente cuando desarrollemos proyectos de investigación.

### **Conceptos de Validez y Confiabilidad:**

Iniciaremos el estudio de este tema, revisando lo que entendemos por error de medición.

La diferencia absoluta entre el valor real de lo que deseamos medir y lo que se obtiene como producto del proceso de medición.

#### **Ejemplo:**

Si deseamos medir la estatura de una persona, aunque no conocemos con certeza la verdadera estatura (estatura real), si ésta es de 1.85 metros y al desarrollar el proceso de medición con un instrumento determinado (un centímetro) obtenemos por ejemplo que la medida de la estatura es de 1.80, entonces, a la diferencia entre el valor real y lo obtenido en el proceso de medición que en este ejemplo es de 0.05 metros es lo que denominamos "error de medición".

El error de medición puede ser minimizado cuando existe una relación directa entre el sistema numérico y el fenómeno que se está midiendo; en este caso, los números representan exactamente las características que se están midiendo y nada más. Obviamente, ésta es una situación idealizada que rara vez se presenta en la práctica.

Generalmente, nuestras medidas tienen algún grado de error, puesto que la escala numérica no representa exactamente el fenómeno bajo investigación. El error total de medición tiene dos componentes. El primero es el error sistemático, el cual origina un permanente sesgo (o desviación) en las mediciones. El segundo es el error aleatorio, éste conlleva influencias que sesgan las mediciones, pero no son sistemáticas.

El error sistemático origina un permanente sesgo en las mediciones, se reproducen constantemente y en el mismo sentido.

Ejemplo: el cero desajustado en una balanza para pesar, de igual manera en un voltímetro para medir voltajes o corrientes eléctricas.

Si consideramos que provocado por un desajuste de medición, una balanza para medir los pesos de determinados objetos tiene por ejemplo un desajuste de 0.5 kg, las mediciones que se realicen en los objetos que estemos interesados en pesar, siempre nos medirán ese desajuste, es así que si el objeto pesa 100kg, la medición que se obtendrá en esa balanza será de 100.5 kg, mostrándonos ese error de medición provocado por el desajuste sistemático.

El error aleatorio, sesga las mediciones pero no es sistemático, es debido a causas irregulares, se produce al azar.



Ejemplo: la presencia de corrientes de aire o variaciones de temperatura, durante la medición, que provoquen alteraciones en dichas mediciones, si requieren precisión, como cuando se están midiendo pesos pequeños de sustancias activas que poseen los medicamentos.

Con la comprensión de los tipos de errores de medición, podemos entonces definir la validez como una medida que se refiere al grado en el cual el proceso de medición esta libre tanto de error sistemático como de error aleatorio.

La confiabilidad, es asociada a una medida que se refiere al grado hasta el cual el proceso de medición esta libre de errores aleatorios.

La confiabilidad se ocupa de la consistencia, exactitud y capacidad de predicción de los hallazgos de la investigación.

La validez está relacionada con la pregunta: ¿estamos midiendo lo que creemos? La validez es un asunto más amplio y difícil que la confiabilidad.

### **En resumen**

Debemos expresar que una medición debe ser confiable para que se le considere válida. En este caso, el error sistemático S y el error aleatorio R, son pequeños o equivalentes a cero. Sin embargo, si una medida no es confiable, no puede ser válida, y si es confiable, entonces puede ser o no ser válida. La confiabilidad es una condición necesaria pero no suficiente para la validez.

Por consiguiente, la validez de una medida es de gran importancia, puesto que involucra tanto el error aleatorio como el sistemático. La facilidad de poder medir la confiabilidad, en comparación con la validez, ha dado mayor énfasis a la presentación de puntajes confiables en lugar de puntajes validos en los estudios de investigación.

Retomemos un ejemplo que nos permite valorar la diferencia entre confiabilidad y validez. Tomado del texto: Diseño de Investigaciones de León y Montero segunda edición Mc. Graw Hill.

"En una gran sala ha habido que acomodar a gran cantidad de estudiantes. Un modo que tiene el profesor para evitar que se copie entre vecinos de pupitre es preparar dos formas distintas para el examen. De este modo, dando ejemplares de examen con formas alternas, se consigue que las preguntas a las que tiene que contestar un determinado alumno no coincidan con las de sus vecinos de derecha e izquierda. Si las dos formas del examen estuvieran bien hechas, lo lógico sería que usted sacara, aproximadamente, la misma nota si realizara los dos exámenes dentro de un intervalo breve de tiempo. Tenga en cuenta que, independientemente de cuál sea su nivel de conocimientos en esa asignatura, lo que sabe en uno y otro momento es lo mismo y, por tanto, una medida fiable (confiable) de sus conocimientos debería dar lugar a los mismos resultados".

Que un procedimiento de medición dé lugar a los mismos resultados habla favorablemente del mismo, en otra forma se dice que es confiable. Sin embargo, no es suficiente garantía para el investigador.

Además de fiable, necesitamos un procedimiento que sea válido. Por validez entendemos que dicho procedimiento recoja los datos que realmente pretende recoger. Sigamos con el ejemplo anterior.

“En la Facultad en la que usted estudia se organizan diferentes grupos dentro del mismo curso. Estos grupos cursan las mismas asignaturas, aunque a veces éstas son impartidas por diferentes profesores. Imagine que todos los profesores de la misma asignatura se ponen de acuerdo para realizar el mismo examen. Imagine también que uno de estos profesores siempre que termina un tema realiza un entrenamiento sistemático del alumno para contestar un determinado tipo de preguntas. Los demás profesores no realizan dicho entrenamiento. Supongamos ahora que usted ha preparado su asignatura junto a un compañero que asiste a la clase del profesor que entrenaba a sus alumnos. Usted tiene la certeza de que su nivel de conocimientos y el de su compañero es muy parecido. Después de realizar el examen, constata con estupor que su nota es un poco superior al aprobado, mientras que la de su compañero es de sobresaliente”. Lo que ha ocurrido es que el examen no es válido para medir los conocimientos de su compañero porque ha mezclado dicha variable con otra. Es decir, en este caso, el examen no mide lo que pretende, nivel de conocimientos, sino otra cosa, eficacia del entrenamiento del profesor añadida al nivel de conocimientos.

### **Reflexiono y ejemplifico:**

Planteo una situación concreta, basándome en la experiencia que me permita relacionar los conceptos de validez y confiabilidad, y me preparo para compartirla con el grupo de clase.

### **En síntesis:**

El proceso en medición nos permite asignar números a los fenómenos que estamos interesados estudiar, este proceso de asignación es imperfecto, lo que nos genera un error de medición. Es de nuestro interés que este error de medición sea el menor posible, para lo cual debemos de tener presente los conceptos de validez y confiabilidad con el fin de identificar si el proceso desarrollado es valido y/o confiable.

Las escalas de medición que estudiamos en Estadística, son: Ordinal, Nominal, Intervalo y Razón, debiendo tener mucho cuidado al establecer la escala de medición que le apliquemos a la variable que nos presenta el fenómeno que estamos estudiando por cuanto al equivocarnos estaremos incrementando el error de medición que para nuestro interés más bien nos conviene minimizarlo.

A continuación atenderemos lo relativo a la medición de actitudes que se vincula de forma esencial a los fenómenos humanos, estudiados por las ciencias sociales, que es a la que pertenecen donde pertenecen la Administración y la Economía.

## **4. Medición de actitudes y escalas de medición de actitudes**

### **¿Qué es un actitud?**

La actitud de un individuo es un conjunto de procesos de percepción permanente y de evaluación, basado en conocimientos y orientado a la acción con respecto a un objeto o fenómeno.

Generalmente, se considera que las actitudes tienen tres componentes principales:

- Un componente cognoscitivo: la percepción que tiene una persona acerca del objeto en cuestión, como su velocidad o durabilidad;
- Un componente afectivo: los sentimientos de una persona hacia el objeto, como "bueno" o "malo" y
- Un componente de comportamiento: la disposición de una persona para responder con su comportamiento al objeto.

En la figura 3 se muestra de forma esquemática la relación de los componentes de las actitudes.

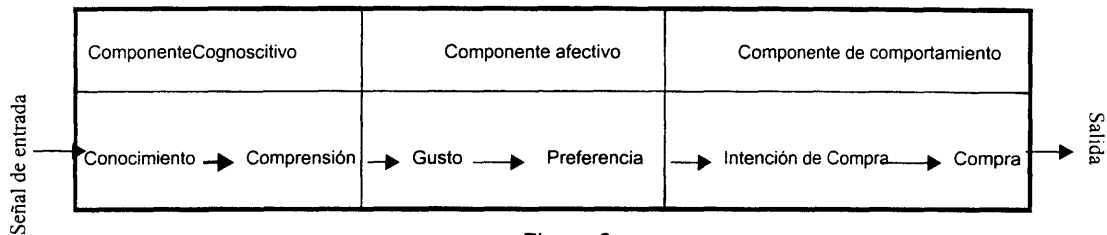


Figura 3

Fuente: Thomas C. Kinnear y James R. Taylor, *Investigación de Mercados. Un enfoque aplicado*. 1994. Página 228.

Si asumiéramos un esquema de sistema, las actitudes pueden descomponerse en los tres componentes mencionados en la figura 3. Ante una señal de entrada al sistema actitud que pudiese ser por ejemplo un mensaje publicitario en algún medio para un producto determinado, se impactan los órganos sensoriales de la persona, lo que permite "conocer" el mensaje y ligarlo a ciertas características que se dicen del producto, lo que conlleva a una comprensión. Cuando esas características del producto concuerdan con la percepción de la persona y aún mas, cuando lo que se le ofrece coincide con su necesidad asume el componente afectivo por medio del "gusto" y se genera la "preferencia" hacia dicho producto, conduciéndolos al componente de comportamiento, haciendo posible una inclinación o intención de compra que conduce a una mayor o menor actitud hacia la compra. La salida es concretamente la manifestación o actitud de respuesta hacia la señal de entrada.

Los tres componentes de la actitud se encuentran en el individuo, en su cerebro, en sus sentimientos y en sus intenciones hacia cómo comportarse ante la señal de entrada. La salida será el comportamiento que manifiesta después de todo un proceso que se genera en él mismo, y que es diferente entre unos y otros seres humanos.

Las actitudes son importantes, entonces, para la toma de decisiones en sistemas humanos, debido a la supuesta relación entre las actitudes y el comportamiento. Normalmente, los modelos que conceptualizan el constructo de la actitud representan una actitud como una serie de componentes secuenciales que generan el comportamiento. La evidencia de la investigación indica que el vínculo entre actitudes y comportamiento no es simplista y que quien toma las decisiones y el investigador debe tener mucho cuidado al suponer que existe tal relación en una situación de decisión.

**Por ejemplo:**

*La predicción del comportamiento futuro para un conjunto de compradores es más factible que la predicción del comportamiento para un comprador individual. Debido a que la mayor parte de las situaciones de decisión están relacionadas con el comportamiento de grupos y no con el comportamiento individual, el vínculo actitud-comportamiento encuentra sustentación empírica en muchas de las situaciones de decisión en los sistemas humanos. Sin embargo, las actitudes son sólo una influencia en el comportamiento, y en una situación de decisión particular otros factores podrían ser más influyentes que las actitudes. Un ejemplo obvio sería el de un individuo que tiene una actitud altamente favorable hacia la compra de un nuevo automóvil pero, debido a las restricciones económicas, tiene que comprar un automóvil usado.*

**a. Componente cognoscitivo.**

El componente cognoscitivo se refiere al conocimiento y la comprensión del encuestado acerca de algún objeto o fenómeno. Algunas veces se conoce con el nombre de componente de creencia. Se expresa por medio de afirmaciones tales como: "Creo que el producto A es..." o "Sé que el producto B podrá...".

El componente cognoscitivo es de gran importancia para muchos tipos de necesidades de información. Muchas situaciones de decisión requieren información acerca de conocimiento / comprensión a manera de ejemplo en investigación de mercado pueden ser las características de un producto, campañas publicitarias, precios, disponibilidad del producto, entre otros.

**b. Componente afectivo.**

El componente afectivo se refiere al gusto y preferencia del encuestado por un objeto o fenómeno. Algunas veces se denomina componente de sentimientos y se expresa por medio de afirmaciones como: "No me gusta el producto A", "La publicidad de X es pobre" y "Prefiero el producto A al producto B".

El componente afectivo es, como el componente cognoscitivo, un aspecto importante de las necesidades de información para muchas situaciones de decisión. Algunos de los ejemplos incluyen la determinación de las preferencias y sentimientos positivos o negativos de los compradores con relación a un producto de nuestra empresa, así como los productos que nos compiten.

**c. Componente de comportamiento.**

El componente de comportamiento se refiere por ejemplo a la intención de compra del comprador y al comportamiento real de compra. La etapa en la cual se presenta la intención de compra se refiere a la predisposición del encuestado para tomar alguna iniciativa de compra antes de tomar la decisión real. Los especialistas en mercadeo se interesan en las intenciones de compra de los encuestados como indicadores del comportamiento futuro de compra.

El comportamiento se refiere a lo que los encuestados han hecho o están haciendo. En mercadeo, por ejemplo el comportamiento se refiere a los patrones de compra y de uso del comprador de un producto o servicio. Generalmente, las necesidades de información se centran en qué, cuánto, dónde y cuándo se realizó la compra, las circunstancias que la rodearon y las características del comprador. La medición del comportamiento incluye la elaboración de una descripción exhaustiva de la situación de compra.

Podemos inducir de lo estudiado hasta ahora, que la tarea de medición en las ciencias del comportamiento usualmente es más difícil y comprende escalas de medición más bajas que las utilizadas en las ciencias físicas.

Un ejemplo crítico de esta dificultad es la medición del constructo de actitud que existe en la mente de los individuos y que no es directamente observable. La escala de actitudes se refiere a las diferentes definiciones operacionales que se han desarrollado para la medición de este constructo.

Al medir actitudes, debemos tener en cuenta los supuestos a nivel de escala y las restricciones que imponen estos supuestos al análisis de datos. Generalmente, las actitudes se miden a un nivel nominal u ordinal, aunque algunos procesos de escalas más complejos permiten mediciones a nivel de intervalos. Siempre existe la tentación de suponer que las mediciones de actitudes tienen las propiedades más poderosas de una escala de intervalo. El investigador siempre debe estar consciente de las siguientes preguntas: ¿Cuáles son las características del constructo que se está midiendo?, ¿Cuáles son las propiedades del sistema numérico que se relacionan adecuadamente con este constructo?

#### **Me pregunto**

¿Por qué compro un producto determinado?, por ejemplo una gaseosa. Relaciono el modelo de los componentes de las actitudes que he estudiado con este producto, y me preparo para compartirlo con mis compañeros.

### **5. La medición de actitudes: Escalas especiales**

En las investigaciones socio-económicas, es muy frecuente la existencia de variables de actitud. Como lo hemos remarcado, son variables más difíciles de medir que las variables físicas (distancia, tamaño, etc), demográficas (edad, sexo) o económicas (renta, ventas, etc).

En general son variables nominales, ordinales y algunos investigadores las consideran como variables de intervalo.

El análisis de estas variables exige frecuentemente un tratamiento especial. Sirve en algunos casos las escalas básicas estudiadas en estadística pero su aplicación obliga a tener en cuenta determinados matices. En otros casos, se suele aplicar escalas especiales, que en concreto son diversos formatos de respuesta ideadas por los investigadores para no limitar o encasillar el número determinado de categorías de respuesta, observado en las escalas de medición básicas.

Algunos tipos de escalas especiales son:

### a. Escala gráfica.

La persona entrevistada debe indicar su posición en un continuo ordenado que va de un extremo de la actitud a otro, que en general se especifica por medio de una línea limitada en los extremos. La línea puede estar numerada en los extremos y / o en puntos intermedios como lo muestra la figura 4.

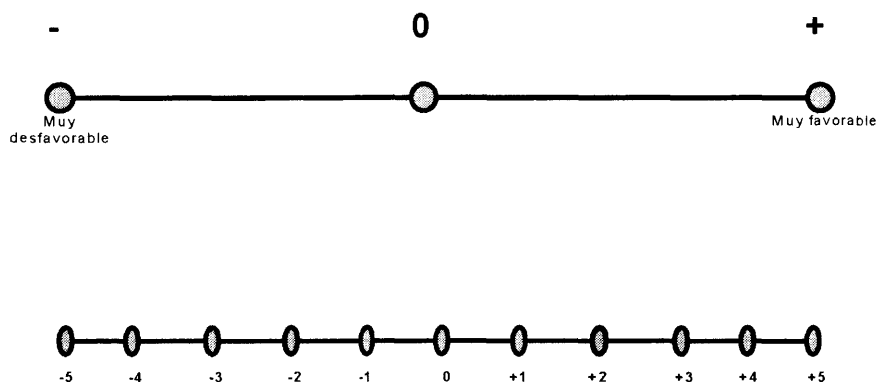


Figura 4

Los extremos pueden ser muy suaves o muy contrastantes. En el primer caso los individuos tenderán a situarse únicamente en los extremos.

Es una escala fácil de construir y de analizar. Si no hay numeración, se mide la distancia hasta el punto base que es el cero.

Existen diversos tipos de escalas gráficas. Una vez que se han registrado los juicios, el investigador puede subdividir el continuo en un conjunto apropiado de categorías y asignar numerales a los juicios.

### b. Escala de categorías

El individuo debe indicar su posición en un número limitado de categorías.



Pueden describirse todas las categorías o solamente las de los extremos. Las descripciones pueden ser mas o menos detalladas.

Esta escala, igual que la escala gráfica, puede utilizarse para muchos temas de investigación socioeconómica.



Número de categorías: En general se considera que el número de categorías debe ser limitado porque si es elevado el entrevistado puede tener dudas en seleccionar entre dos o más de ellas. Un número recomendado es el incluido entre 5 y 7. En algún caso el número adecuado puede ser de dos, cuando se trata de responder si o no.

Número impar o par: en el primer caso hay una posición neutral, mientras que en el segundo caso no existe.

Escala balanceada o no balanceada. En el primer caso el número de categorías favorables coincide con el de categorías no favorables. Es el caso más general, aunque la escala no balanceada puede utilizarse cuando se piensa que las actitudes se concentran en una dirección más que en la otra.

Numeración de las categorías. Útil para informatizar y sobre todo si se trata de una escala de intervalos, es decir cuando rige el principio de igualdad de diferencias.

Individuos que no son capaces de responder. Cuando se piensa que hay un porcentaje importante de individuos que no pueden o no desean contestar, con el objeto de no forzarlo, es útil incluir una categoría que suele denominarse "no sabe, no contesta". Si la escala es de intervalos, habrá que tener en cuenta su numeración con fines de calcular valores medios y varianzas.

### **c. Escala de orden jerárquico**

Cada entrevistado debe ordenar un conjunto de objetos o hechos según su preferencia o según una determinada característica (sabor, calidad, etc). Implica una selección forzada. Es difícil su aplicación para un número elevado de objetos.

Es de carácter comparativo. Los defensores de esta técnica consideran que se trata de un proceso similar al de la decisión (compra u otros comportamientos).

La técnica es muy sencilla y muy usada. En general se procede a sumar las respuestas. El resultado es una escala ordinal.

producto	respuestas	suma
A	23123	11
B	11211	6
C	32332	13 (B más preferido)

Si la escala ubica mayor preferencia con números menores, si sucede lo contrario, el más preferido será el C.

En el grupo de clase o en su trabajo levante una encuesta de 10 personas utilizando un formato de respuesta de orden jerárquico que nos permita comparar por ejemplo la preferencia de cuatro bebidas gaseosas en base a una característica de sabor que puede ser: nivel de dulzura. Prepárese a compartir los resultados con el grupo.

#### d. Escala de comparación por pares

Con esta técnica, se presenta a los encuestados dos objetos de un conjunto y se les solicita que seleccionen uno con relación a la actitud estudiada.

Si hay  $n$  objetos, el número de comparaciones es de  $\frac{n(n-1)}{2}$ .

Para  $n = 5$ , el número de comparaciones es de 10. El número de comparaciones aumenta sensiblemente cuando aumenta  $n$ . Así para  $n = 10$ , el número de comparaciones pasa a 45.

Si llamamos 1 a la preferencia de un objeto columna sobre el otro objeto fila del par y 0 a la no preferencia, puede establecerse el siguiente cuadro, en el supuesto de  $n = 5$ .

En general, el número 1 corresponde a un número de respuestas favorables, superior al 50 por ciento de los encuestados.

	A	B	C	D	E	
A	*	1	1	0	0	A partir del total se obtiene una escala ordinal
B	0	*	0	0	0	
C	0	1	*	0	0	
D	1	1	1	*	1	B>C>A>E>0 (B más preferido)
E	1	1	1	0	*	
Total	2	4	3	0	1	

Como desventaja, se cita que el proceso de comparación por pares no se corresponde con el proceso real de elección y que el orden de los pares puede influir sobre los resultados. Un proceso más complejo permite obtener una escala de intervalos.

#### Me ejercito:

Preparo una escala de comparación por pares para comparar cuatro marcas de pizzas: Hut, Valenti's, House, Domino's. Encuesto a 5 personas del grupo de clase o del trabajo y me preparo a compartir los resultados con el grupo.

#### e. Escala de suma constante

Consiste en distribuir un número determinado de puntos, por ejemplo 100 puntos entre un cierto número de objetos, también en base a preferencias o a una determinada característica.

Si a un objeto A se le da doble puntuación que a un objeto B, se supone que se le considera un valor doble. La técnica es útil cuando el número de objetos es pequeño.

#### Para comprender mejor realizaremos el siguiente ejemplo:

Distribuya 100 puntos para las cuatro asignaturas siguientes que usted ha estudiado: Introducción a la Administración, Economía Política, Matemáticas para Administración I y Estadística. Asigne mayor puntaje a la asignatura de menor preferencia y menos puntaje a la asignatura de menor preferencia, recuerde que la suma de la asignación de puntos para las cuatro asignaturas debe sumar 100 puntos. Comparta con sus compañeros los resultados.

## f. Escala semántica diferencial

Los encuestados deben clasificar cada objeto según varias escalas en cuyos extremos hay dos adjetivos. Cada escala corresponde a un atributo o a una característica.

En cada escala existen 5 a 7 respuestas

El número de escalas puede llegar a 20 ó 25.

Si las escalas tienen una graduación de positivo a negativo, se suele rotar la colocación del aspecto positivo a la derecha o a la izquierda, para evitar respuestas sistemáticas.

*Un ejemplo:*

Moderno	_____	_____	_____	_____	_____	Anticuado
Rápido	_____	_____	_____	_____	_____	Lento
Alegre	_____	_____	_____	_____	_____	Triste
Precio alto	_____	_____	_____	_____	_____	Precio bajo

Las respuestas vienen numeradas de 1 a 5 o de -2 a +2 y las escalas vienen consideradas generalmente como de intervalos (tema dudoso como en la mayor parte de las escalas de intervalo).

Los resultados pueden presentarse en forma de perfil:

-	A	-	B	-	-
-	-	A	-	B	-
-	-	A	B	-	-

Los resultados numéricos de las distintas escalas, también pueden sumarse obteniendo así una clasificación numérica única, aunque si se utiliza esta última como resultado único se pierde información.

Es ampliamente utilizada en mercadotecnia para estudiar el posicionamiento de productos, imagen, marcas y empresas.

## g. Escala de Stapel

Muy similar a la escala semántica diferencial, es una modificación de la escala de diferencial semántico. La diferencia es que cada escala se basa en un solo adjetivo, existiendo 10 respuestas en lugar de las 7 más habituales de la escala semántica.

+5	+4	+3	+2	+1	-1	-2	-3	-4	-5
Buen Servicio									

La escala está diseñada para medir simultáneamente la dirección e intensidad de las actitudes. Se diferencia del diferencial semántico en que los valores de la escala indican hasta qué punto la descripción o el adjetivo se ajusta al objetivo evaluado.

## h. Escala de Likert

En lugar de utilizar escalas basadas en adjetivos opuestos, se utilizan frases o enunciados. El encuestado debe indicar su grado de aceptación o discrepancia. Las respuestas están valoradas de 1 a 5 o de -2 a +2

**Por ejemplo:**

Muy de acuerdo	+2	5
De acuerdo	+1	4
Neutral	0	3
En desacuerdo	-1	2
Muy en desacuerdo	-2	1

Las preguntas deben ser claramente favorables o desfavorables al tema en cuestión, puntuándose las desfavorables en sentido contrario a las favorables.

Una vez hechas las entrevistas, es necesario eliminar las escalas en las que la mayor parte de los encuestados han contestado lo mismo.

También hay que eliminar las escalas que no discriminan entre los encuestados que tienen una actitud favorable y los que tienen una actitud desfavorable; por ejemplo los enunciados que generan respuestas distribuidas aleatoriamente.

Para cada encuestado se calcula la suma. Esta es una técnica utilizada para dividir un cierto número de individuos en dos o más grupos. Como en la escala semántica, se supone que se está trabajando con escalas de intervalos.

***A manera de ejemplo, valoremos la siguiente pregunta con escala Likert:***

*¿El primer poder de la República es el Legislativo? Las opciones de respuesta se ubicarán como:*

*Muy de acuerdo:*

*De acuerdo:*

*Ni de acuerdo/ Ni en desacuerdo:*

*En desacuerdo:*

*Muy en desacuerdo:*

### **Sinteticemos:**

La actitud en los humanos es de importancia estudiarla para tomar decisiones. Hemos identificado que las actitudes forman un complejo sistema que contiene componentes como el cognoscitivo, el afectivo y el comportamental. Las actitudes se pueden medir por medio de los niveles de escala de medición estudiados en estadística, pero además, existe una diversidad de escalas especiales que nos pueden permitir la definición del formato de respuesta para la investigación de las actitudes.

En este primer tema pusimos énfasis en cómo la realidad impacta en nuestros sentidos, cómo estas señales son percibidas por nosotros y hacen posible la conexión con nuestro cerebro, para poder abstraer ese pedazo de realidad y explicarla en forma de conocimiento.

Hemos identificado la importancia que debemos darle a los datos para poder desentrañar la información que subyace en la realidad. Para poder obtener los datos adecuados debemos realizar mediciones de los fenómenos que percibimos, existiendo una verdadera

dificultad por el hecho de realizar una asociación entre el sistema empírico que existe en el mundo real que nos rodea, con el sistema abstracto que nos permite efectuar las mediciones sobre la base del sistema numérico.

Esa asociación, debemos de identificarla claramente en las variables que estamos interesados en medir, en especificar su tipo, si son cuantitativas o cualitativas, si son cuantitativas discretas o continuas; adicionalmente, que tipo de escala de medición podemos aplicarles sean nominal, ordinal, de intervalo o de razón.

Todo lo anterior nos obligó a revisar lo que estudiamos en el módulo autoformativo de estadística, dado que nos servirá para realizar el plan de trabajo final.

### **Actividad de Autoaprendizaje No. 1:**

1. Justifico la importancia del proceso de medición
2. Respondo lo siguiente:
  - a. ¿A qué se le denomina variable de actitud? ¿Cuáles son sus componentes?
  - b. ¿Qué diferencias existen entre las escalas de Likert y la Semántica Diferencial?
3. Explico el proceso que sigue la escala de comparación por pares
4. Anoto ejemplos de error sistemático y de error aleatorio.

Luego de realizar esta actividad de autoaprendizaje, busco las respuestas en las páginas 113 y 114 al final de la unidad autoformativa I y me corrijo.

## **B. POBLACIÓN Y MUESTREO. TIPOS DE MUESTREO:**

### **1. Concepto de población y muestreo**

En distintas situaciones en nuestras actividades nos encontramos ante la realidad del muestreo, tal es el caso cuando nos sacamos una muestra de sangre en el laboratorio o cuando se realiza en la universidad, en nuestro grupo de clase el levantamiento de una muestra para evaluar a nuestro profesor.

El origen del muestreo está en la teoría de las probabilidades que ya ha sido estudiado por nosotros en el módulo de estadística. En este módulo estudiamos una cantidad de conceptos que nos servirán para el estudio del muestreo, principalmente lo relacionado con la teoría de las probabilidades, la inferencia estadística, las distribuciones muestrales, así como otros conceptos tales como media, desviación estándar, varianza, entre otros.

El conocimiento de los conceptos básicos de Estadística es pues un requisito para el estudio de los diseños de encuestas por muestreo. Es por eso que es conveniente que con nuestro material autoformativo de estadística, revisemos estos conceptos básicos. Sin embargo conviene que aquí complementemos algunos conceptos:

#### **a. Concepto de Población:**

##### ***Población la defino como:***

- *Conjunto o cúmulo de individuos u objetos cuyas propiedades se han de analizar.*
- *Conjunto de todos los posibles individuos, personas, objetos o mediciones de interés estadístico.*

Una población puede estar formada por personas como todos los estudiantes inscritos en una universidad, todos los alumnos de una clase de Técnicas de Recopilación de Información, o todos los trabajadores de la empresa donde laboro.

Una población también puede estar formada por objetos, como las llantas producidas durante una semana en una fábrica, o todos los vehículos que existen en Managua.

Una población también puede estar formada por un grupo de medidas, como podrían ser los pesos de las barras de mantequilla que se producen por una empresa de productos lácteos o el volumen del contenido de jarabe para la tos que produce y vende una empresa farmacéutica.

*Como usted puede observar, una población en sentido estadístico no necesariamente se refiere a personas.*

Una población está determinada por sus características definitorias. Por lo tanto, el conjunto de elementos que posea esta característica se denomina población o universo. Población es entonces la totalidad del fenómeno a estudiar, donde las unidades de población poseen una característica común, la que se estudia y da origen a los datos de la investigación.



Entonces, una población es el conjunto de todas las cosas que concuerdan con una serie determinada de especificaciones. Un censo, por ejemplo, es el recuento de todos los elementos de una población; en nuestro país el último que se le aplicó a toda la población, se realizó en 1995.

Cuando seleccionamos algunos elementos y no todos, con la intención de averiguar algo sobre una población determinada, nos referimos a este grupo de elementos como muestra. Por supuesto, esperamos que lo que averiguamos en la muestra sea cierto para la población en su conjunto. La exactitud de la información recolectada depende en buena medida de la forma en que fue seleccionada la muestra.

Cuando no es posible medir cada uno de los individuos de una población, se toma una muestra representativa de la misma.

La muestra descansa en el principio de que las partes representan al todo y, por lo tanto, refleja las características que definen la población de la que fue extraída, lo cual nos indica que es representativa. Por lo tanto, la validez de la generalización depende de la validez y el tamaño de la muestra.

### **Diferencia entre muestras y poblaciones:**

Muestra y población son términos relativos. Una población es un todo y una muestra es una fracción o segmento de ese todo. Estudiamos muestras con el fin de ser capaces de describir poblaciones.

### **b. Concepto de muestra**

#### ***Una muestra se define como:***

- |  |
|--|
| <ul style="list-style-type: none"><li>▪ <i>Un subconjunto de una población.</i></li><li>▪ <i>Una porción, o parte de una población de interés.</i></li></ul> |
|--|

El estudio de muestras es más sencillo que el estudio de la población completa, cuesta menos y lleva menos tiempo. Además, se ha probado que el examen de una población entera todavía permite la aceptación de elementos defectuosos, por tanto, en algunos casos, el muestreo puede elevar el nivel de calidad de la información que recabamos.

Una población es un conjunto de todos los elementos que estamos estudiando, acerca de los cuales intentamos sacar conclusiones. Debemos definir dicha población de modo que quede claro cuándo un cierto elemento pertenece o no a la población.

Una muestra es una colección de algunos elementos de la población, pero no de todos. Cualquier grupo que cumple con los requisitos de la población, puede constituir una muestra, siempre y cuando el grupo sea una fracción de la población completa.

Una muestra representativa contiene las características relevantes de la población en las mismas proporciones en que están incluidas en tal población.

### c. Leyes del método de muestreo:

El método de muestreo se basa en ciertas leyes que le otorgan su fundamento científico, y son:

- 1) **Ley de los grandes números:** si en una prueba, la probabilidad de un acontecimiento o suceso es  $P$ , y si éste se repite una gran cantidad de veces, la relación entre las veces que se produce el suceso y la cantidad total de pruebas (es decir, la frecuencia  $F$  del suceso) tiende a acercarse cada vez más a la probabilidad  $P$ .
- 2) **Cálculo de probabilidades:** La probabilidad de un hecho o suceso es la relación entre el número de casos favorables ( $p$ ) a este hecho con la cantidad de casos posibles, suponiendo que todos los casos son igualmente posibles. El método de establecer la probabilidad es lo que se denomina cálculo de probabilidad.

De estas dos leyes fundamentales de la estadística, se infieren aquellas que sirven de base más directamente al método de muestreo:

- 3) **Ley de la regularidad estadística:** Un conjunto de  $n$  unidades tomadas al azar de un conjunto  $N$ , es casi seguro que tenga las características del grupo más grande.
- 4) **Ley de la inercia de los grandes números:** Esta ley es contraria a la anterior. Se refiere al hecho de que en la mayoría de los fenómenos, cuando una parte varía en una dirección, es probable que una parte igual del mismo grupo, varíe en dirección opuesta.
- 5) **Ley de la permanencia de los números pequeños:** Si una muestra suficientemente grande es representativa de la población, una segunda muestra de igual magnitud deberá ser semejante a la primera; y, si en la primera muestra se encuentran pocos individuos con características raras, es de esperar encontrar igual proporción en la segunda muestra.

Sin embargo, no todas las muestras resultan útiles para llevar a cabo un trabajo de investigación. Lo que se busca al emplear una muestra es que, observando una porción relativamente reducida de unidades, se obtengan conclusiones semejantes a las que lograríamos si estudiáramos el universo total. Cuando una muestra cumple con esta condición, es decir, cuando nos refleja en sus unidades lo que ocurre en el universo, la llamamos muestra representativa. Por lo tanto, una muestra representativa contiene las características relevantes de la población en las mismas proporciones en que están incluidas en tal población. Sus conclusiones son susceptibles de ser generalizadas al conjunto del universo, aunque para ello debemos añadir un cierto margen de error en nuestras proyecciones.

#### ***¿Por qué tomar una muestra en vez de estudiar cada integrante o elemento de una población?***

Una muestra de votantes registrados es necesaria debido al costo prohibitivo de comunicarse con millones de electores antes de efectuar una elección, esto es lo que diversas firmas encuestadoras han realizado en Nicaragua previo a las elecciones municipales y nacional realizadas años recientes en nuestro país.

Al someter a una prueba de contenido de humedad, a un grano de trigo, se le destruye; por lo tanto, el tomar la muestra de toda una cosecha en vez de toda la cosecha de esta gramínea resulta imperativo. Si los catadores de ron consumieran todo éste en las pruebas, nada quedaría para la venta. Sería físicamente imposible para un reducido grupo de biólogos marinos, capturar y marcar todas las focas del océano. Estas y otras razones nos obligan a estudiar la teoría de muestreo.

#### **d. ¿Qué es el Muestreo?**

*En esencia el muestreo consiste en obtener información acerca de un amplio grupo o "universo", valiéndose de una parte representativa del mismo.*

Este universo representa al total acerca del cual se desea conocer algún dato o datos.

Las circunstancias de un problema en particular determinan el método o combinación de métodos de muestreo que se ha de emplear.

Podemos identificar algunas ventajas importantes en la aplicación del muestreo:

##### **1) Ventajas del Muestreo.**

- Nuestros conocimientos, actitudes y acciones están basados en gran parte sobre muestras. Esto es verdad real en la vida cotidiana como en la investigación científica. Tanto en la ciencia como en los asuntos humanos careceremos de los recursos necesarios para estudiar más de un fragmento de los fenómenos que pueden ampliar nuestros conocimientos.
- En la mayoría de las aplicaciones para las que se creó esta teoría, el conjunto del que se desea obtener información es finito y bien delimitado como sería por ejemplo el conjunto de habitantes de una ciudad, las máquinas de una fábrica o los peces de un lago. En ciertos casos, puede parecer factible obtener la información por medio de una enumeración completa o censo del conjunto.
- Las ventajas del muestreo en comparación con la enumeración completa, están asociadas a.
- Costo reducido. Si los datos se obtienen únicamente de una pequeña fracción del total, los gastos son menores que los que se realizarían si se llevara a cabo un censo completo. En poblaciones muy grandes se pueden obtener resultados lo suficientemente precisos cuando se analizan muestras que representan sólo una pequeña fracción de la población.

Las encuestas realizadas para obtener información relacionadas con las ventas o campañas de publicidad en la investigación de mercados, pueden emplear muestras de sólo unos cuantos cientos o miles de observaciones.

- Mayor rapidez. Por la misma razón, los datos pueden ser recolectados y resumidos más rápidamente con una muestra que con una enumeración completa. Esta es una consideración vital cuando se necesita la información con urgencia.
- Más posibilidades. Para obtener la información en ciertos tipos de encuestas, se utilizan los servicios de personal altamente calificado o equipo muy especializado de disponibilidad limitada. Por lo tanto, en estos casos el censo completo es impracticable y como alternativa a la obtención de datos por muestreo, sólo existe la de no obtenerlos. De ahí que las encuestas basadas en el muestreo tienen más posibilidades y flexibilidad respecto a la información que puede obtenerse.
- Mayor precisión. Debido a que al reducir el volumen de trabajo se puede emplear personal más capacitado y someterlo a un entrenamiento intensivo y debido también a que en estas condiciones será factible la supervisión cuidadosa del trabajo de campo y del procesamiento de los resultados, una muestra puede producir resultados más precisos que la enumeración completa.

Sin embargo, no necesariamente en todos los casos de estudio nos encontramos con ventajas del muestreo; en algunas situaciones nos podemos encontrar con limitaciones.

## **2) Limitaciones del Muestreo**

- El muestreo no es factible en toda situación en que se requieren conocimientos acerca de cada una de las unidades componentes de un universo estadístico.
- Si bien muchas veces los incrédulos exageran estos aspectos, lo cierto es que de vez en cuando aparece alguna muestra "anómala" que arroja resultados falsos.
- El error de muestreo puede ser más grande de lo que se esperaba si la muestra se planifica o se realiza mal.
- Además en muchas ocasiones, las unidades que se han de medir son tan raras y su variabilidad es tan grande, que hacer una muestra sería un derroche de recursos y de tiempo.

## **e. Muestreo y Estadística**

El muestreo es sólo uno de los componentes, sin duda uno de los más importante, de una amplia disciplina de métodos científicos que se conoce con el nombre de Estadística.

La Estadística proporciona los medios para reunir, analizar e interpretar las engorrosas cantidades de datos que la mayoría de las empresas deben procesar en esta era científica. En muchos casos, el aluvión de datos no corresponde a todo el universo o población que se desea estudiar y, en consecuencia, como muchas de las cifras con que contamos se obtienen de las muestras, no cabe duda de que convendrá dedicar cierta atención a los diversos métodos que existen para seleccionar las muestras y para calcular cosas como promedios, proporciones y totales a partir de las muestras.

## **f. Muestreo y Probabilidad**

Al hacer un cálculo sobre la base de los datos de una muestra, siempre se corre el riesgo que la muestra en cuestión no represente bien al universo o población que se investiga. Este riesgo se reduce a un mínimo empleando métodos de "probabilidad" y técnicas estimativas apropiadas y tomando una muestra de suficiente magnitud, aunque no por ello el riesgo desaparece del todo.

## **g. Error de Muestreo**

Siempre que calculamos algo a partir de una muestra, es muy importante que el cálculo sea exactamente igual a la cifra que se obtiene realizando la investigación en el 100 por ciento de los sujetos involucrados. La diferencia entre lo obtenido de la muestra y lo que deberíamos obtener del total de la población es el error de muestreo, ello siempre que los datos de la muestra y los datos de la población completa se reúnan empleando métodos idénticos.

## **h. Exactitud y Precisión**

En Estadística se acostumbra aludir a la precisión y no a la exactitud de los cálculos basados en muestras. La exactitud de un cálculo es el grado en que éste se aproxima a la cifra real, mientras que la precisión refleja el grado en que se aproxima a la cifra que se obtendría con un análisis del 100 por ciento, si se empleasen métodos idénticos de recolección de datos.

## **2. Tipos de Muestras o Estrategias de muestreo**

Es indudable que toda muestra entraña algún error. Sin embargo, para ciertos tipos de muestra, el grado de precisión que se desea en el cálculo se puede estipular de antemano. Estas muestras se denominan muestras de probabilidad o científicas y se caracterizan por el hecho de que se conoce la probabilidad de selección de cada unidad.

Por lo tanto, las muestras pueden ser clasificadas, en una primera división en probabilísticas y no probabilísticas como lo muestra la figura 5.

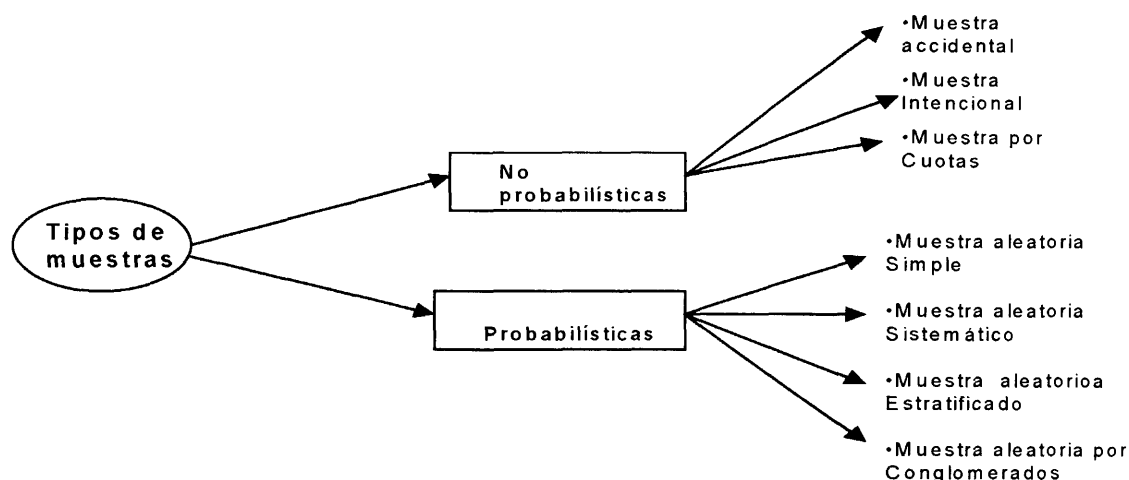


Figura 5

En las muestras probabilísticas, la característica fundamental es que todo elemento del universo tiene una determinada probabilidad de integrar la muestra, y esa probabilidad puede ser calculada matemáticamente con precisión. En las muestras no probabilísticas ocurre lo contrario y el investigador no tiene idea del error que puede estar introduciendo en sus apreciaciones.

### a. Muestras No Probabilísticas

Las muestras no probabilísticas más usadas son:

- 1) **Muestra accidental.** Es aquella que se obtiene sin ningún plan preconcebido; las unidades elegidas resultan producto de circunstancias fortuitas. Si entrevistamos a los primeros 50 transeúntes que pasan por cierta calle o medimos la profundidad del mar a lo largo de un trayecto entre dos puntos cualesquiera, estaremos en presencia de una muestra accidental; los datos obtenidos podrán o no representar al universo en estudio. El investigador no puede saber hasta qué punto sus resultados podrán proyectarse, con confiabilidad, hacia el conjunto más amplio que desea conocer.
- 2) **Muestra intencional.** Las unidades se eligen en forma arbitraria, designando a cada unidad según características que para el investigador resulten de relevancia. Se emplea, por lo tanto, el conocimiento y la opinión personal para identificar aquellos elementos que deben ser incluidos en la muestra. Se basa, primordialmente, en la experiencia de alguien con la población. Estas muestras son muy útiles y se emplean frecuentemente en los estudios de caso, por más que se quiera, la posibilidad de generalizar conclusiones a partir de ellas, será en rigor nula. En algunas oportunidades se usan como guía o muestra tentativa para decidir cómo tomar una muestra aleatoria más adelante.
- 3) **Muestra por cuotas.** Consiste en predeterminar la cantidad de elementos de cada categoría que habrán de integrar la muestra para lo cual se divide a la población en estratos o categorías, y se asigna una cuota para las diferentes

categorías y, a juicio del investigador, se selecciona las unidades de muestreo. La muestra debe ser proporcional a la población, y en ella deberán tenerse en cuenta las diferentes categorías.

Así podemos asignar una cuota de 50 hombres y 50 mujeres a una muestra de 100 individuos, asumiendo que ésta es la distribución de la población total. Por más que esa presunción llegue a ser válida, no deja de existir cierta arbitrariedad en este modo de proceder, por lo que la rigurosidad estadística de las muestras por cuotas se reduce considerablemente.

El muestreo por cuotas se presta a distorsiones, al quedar a criterio del investigador la selección de las categorías.

## b. Muestras Probabilísticas

Recordando, en ellas cada uno de los elementos del universo tiene una probabilidad determinada y conocida de ser seleccionado. Los procedimientos más usuales para la obtención de muestras aleatorias son:

- 1) **Muestreo aleatorio simple.** Selecciona muestras mediante métodos que permiten que cada posible muestra tenga igual probabilidad de ser seleccionada y que cada elemento de la población total tenga una oportunidad igual de ser incluido en la muestra.

La forma más común de obtener una muestra aleatoria es la selección al azar. Es decir, cada uno de los individuos de una población tiene la misma posibilidad de ser elegido. Si no se cumple este requisito, se dice que la muestra es viciada. Para tener la seguridad de que la muestra aleatoria no es viciada, debe emplearse para su constitución una tabla de números aleatorios o un programa de computadora para la elección de números al azar.

- 2) **Muestreo aleatorio sistemático.** En el muestreo sistemático, los elementos son seleccionados de la población dentro de un intervalo uniforme que se mide con respecto al tiempo, al orden o al espacio.

El muestreo sistemático difiere del aleatorio simple en que cada elemento dentro de un intervalo uniforme tiene igual probabilidad de ser seleccionado, pero cada muestra no tiene una posibilidad igual de ser seleccionada.

- 3) **Muestreo estratificado.** Una muestra es estratificada cuando los elementos de la muestra son proporcionales a su presencia en la población. La presencia de un elemento en un estrato excluye su presencia en otro. Para este tipo de muestreo, se divide a la población en varios grupos o estratos con el fin de dar representatividad a los distintos factores que integran el universo de estudio. Para la selección de los elementos o unidades representantes, se utiliza el método de muestreo aleatorio.
- 4) **Muestreo por conglomerados o por racimo.** Dividimos la población en grupos, o conglomerados, y luego seleccionamos una muestra aleatoria de estos conglomerados. Suponemos que estos conglomerados individualmente



son representativos de la población como un todo (Por ejemplo: las cuadras o barrios de un pueblo). Un procedimiento de conglomerado bien diseñado puede producir una muestra más precisa a un costo considerablemente menor que el de un muestreo aleatorio simple.

Tanto en el muestreo estratificado como en el de conglomerado, la población se divide en grupos bien definidos. Usamos el muestreo estratificado cuando cada grupo tiene una pequeña variación dentro de sí mismo, pero hay una amplia variación dentro de los grupos. Usamos el muestreo de conglomerados en el caso opuesto, cuando hay una variación considerable dentro de cada grupo, pero los grupos son esencialmente similares entre sí.

En el caso de las muestras probabilísticas, estas deben de cumplir algunas condiciones que son:

- Homogeneidad: debe ser extraída de la misma población.
- Independencia: las observaciones no deben estar mutuamente condicionadas entre sí.
- Representatividad: la muestra debe ser el mejor reflejo posible del conjunto del cual proviene.

5) **Otros tipos de muestras.** Estas se refieren a una combinación entre muestras probabilísticas y no probabilísticas, una de ellas es:

Muestreo mixto: se combinan diversos tipos de muestreo. Por ejemplo: se puede seleccionar las unidades de la muestra en forma aleatoria y después aplicar el muestreo por cuotas.

Conviene tener claro que muchas veces no existe el muestreo óptimo y serán aspectos de tipo económico o de tiempo los que determinen su elección. Sin embargo, siempre se ha de tener presente el tipo de muestreo y el tamaño de la muestra para la inferencia.

A manera de síntesis, hemos estudiado el concepto de población y de muestra, en el caso de la población lo relacionamos a la globalidad o el conjunto universo de elementos que estamos interesados en estudiar. La muestra, mientras tanto es un subconjunto de esa población que dadas nuestras limitantes nos permiten estudiar los elementos contenidos en ellas y a partir de los resultados estadísticos que nos proporcionen nos permitirán realizar inferencias sobre la población.

Hemos diferenciado en base al grado de precisión que se desee, estipular de antemano en cuanto al error, las muestras probabilísticas y no probabilísticas, así como algunos tipos de muestras para cada una de ellas. Convendrá ahora atender lo que denominaremos los métodos de recopilación de los datos, esto nos permitirá atender el "cómo" recopilamos los datos necesarios para nuestro estudio.

## Actividad de Autoaprendizaje No. 2

1. Elaboro un cuadro comparativo en el que se observen los distintos tipos de muestras estudiadas.
2. Busco un ejemplo en un diario u otro medio de información sobre alguna muestra aplicada en algún caso de interés social. Analizo y escribo mis observaciones en mi texto paralelo sobre el tipo de muestra, conclusiones, validez, representatividad.

Luego podré buscar las respuestas a esta actividad en las páginas No. 115 y 116, de la hoja de respuesta que aparece al final de la unidad autoformativa I y me retroalimenta.

## C. MÉTODOS DE RECOPIACIÓN DE DATOS<sup>1</sup>

Denominamos método de recopilación, al procedimiento que seguimos para obtener los datos o información de la fuente donde se producen y que estamos interesados en obtener.

Los dos métodos básicos de recolección de datos de los encuestados son la comunicación y la observación como se puede contemplar en la figura 6.

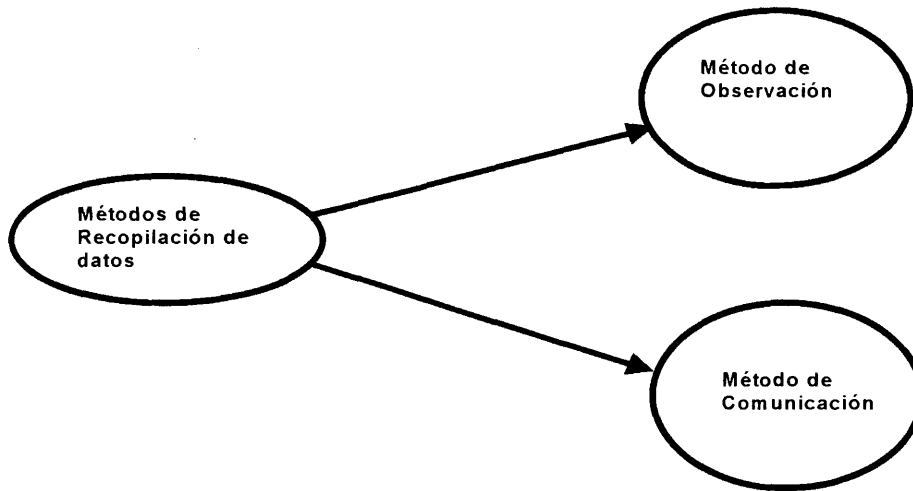


Figura 6

### 1. Método de observación

Se utiliza para recolectar los datos necesarios para un estudio. La observación es un método clásico de recolección de datos en la investigación científica; además, es la manera básica por medio de la cual obtenemos información acerca del mundo que nos rodea.

Los Principios básicos para realizar una observación se asocian a:

- Debe tener un propósito específico.
- Debe ser planeada cuidadosa y sistemáticamente.
- Debe llevarse, por escrito, un control cuidadoso de la misma.
- Debe especificarse su duración y frecuencia.
- Debe seguir los principios básicos de confiabilidad y validez.

La observación abarca el registro del comportamiento del encuestado, es el proceso de reconocimiento y registro del comportamiento de las personas, objetos y eventos.

<sup>1</sup> Parte del presente tema ha sido tomado de Thomas C. Kinneary y James R. Taylor, 1994. Investigación de Mercados. Un enfoque aplicado. Cuarta edición. Editorial Mc. Graw Hill, Colombia. Capítulo X, página 279-305.

Quienes toman las decisiones en la empresa, utilizan extensamente la observación informal. Los especialistas de mercadeo por ejemplo toman nota de los patrones de compra de los clientes, de la publicidad y precios de la competencia, de la disponibilidad del producto, etc. El peligro de sacar conclusiones de la observación informal, como en el caso de la comunicación informal, es que el potencial de los errores muestrales y de los no muestrales es muy grande. Por consiguiente, se han diseñado técnicas para la observación formal con el fin de controlar estos errores y proporcionar datos válidos para la toma de decisiones.

Rara vez un diseño de investigación en administración se basa completamente en el método de observación. En la práctica las técnicas de observación se utilizan conjuntamente con otras técnicas de recolección de datos. Es importante establecer las ventajas y las desventajas del método de observación para identificar su papel en el conjunto de instrumentos de recolección de datos que tiene a su disposición el investigador.

#### **a. Ventajas del método de observación.**

El método de observación presenta varias ventajas cuando se compara con el método de comunicación. En primer lugar, no se basa en la buena voluntad del encuestado para suministrar los datos deseados. En segundo lugar, se reduce o elimina el sesgo potencial causado por el entrevistador y el proceso de entrevista. Por tanto los datos de observación deben ser más exactos. En tercer lugar, algunos tipos de datos sólo pueden recolectarse mediante la observación. Obviamente aquellos patrones de comportamiento de los que el encuestado no tenga conciencia pueden registrarse sólo mediante la observación.

#### **b. Desventajas del método de observación.**

El método de observación tiene dos grandes deficiencias que limitan su utilización en forma significativa. Primero, la incapacidad para observar aspectos tales como el conocimiento, las persuasiones, los sentimientos y las preferencias. Además es difícil observar un sinnúmero de actividades personales e íntimas, como aplicarse maquillaje y desodorante, comer, jugar familiarmente con los niños y ver televisión en horas avanzadas de la noche. Segundo, los patrones de comportamiento observados deben ser de corta duración, ocurrir con frecuencia o ser razonablemente predecibles si es que realmente los requisitos de costos y tiempo para la recolección de datos son competitivos con otras técnicas de recolección de datos. Este requisito limita el método de observación a un conjunto único de circunstancias.

#### **c. Clasificación de las técnicas de observación.**

Las técnicas de observación se pueden clasificar de cinco maneras:

- 1) Observación natural o artificial
- 2) Observación oculta o no oculta

- 3) Observación estructurada o no estructurada
- 4) Observación directa o indirecta y
- 5) Observación humana o mecánica.

Generalmente las técnicas de observación presentan ciertos grados de estas características en vez de la distinción dicotómica que se presentó anteriormente.

- 1) **Observación natural versus observación artificial.** La observación natural abarca la observación del comportamiento tal como se presenta normalmente en el medio ambiente, por ejemplo, hacer compras en un supermercado. La observación artificial comprende la creación de un ambiente artificial y la observación de los patrones de comportamiento que presentan las personas situadas en este medio ambiente, por ejemplo, hacer que las personas compren en un supermercado simulado.

La ventaja de un medio ambiente más natural es que existe una mayor posibilidad de que el comportamiento exhibido refleje con mayor precisión los patrones reales de comportamiento. En contraste esto incrementa los costos agregados por tener que esperar a que suceda el comportamiento y la dificultad para medir el comportamiento en un ambiente natural.

- 2) **Observación oculta versus observación no oculta.** El ocultamiento se refiere al hecho de que los encuestados estén o no conscientes de que se les está observando. El papel del observador debe ocultarse en situaciones en las cuales las personas se comportarían de manera diferente si saben que están observándolas. Pueden emplearse diversos enfoques como espejos de doble faz, cámaras escondidas y observadores vestidos como vendedores para ocultar la observación.
- 3) **Observación estructurada versus observación no estructurada.** La observación estructurada es apropiada cuando el problema de investigación se ha definido claramente y la especificación de las necesidades de información permite una identificación precisa de los patrones de comportamiento que deben observarse y medirse. La observación no estructurada es apropiada en situaciones en las cuales todavía no se ha formulado el problema de investigación y se necesita una gran flexibilidad en la observación para desarrollar hipótesis que sean útiles para definir el problema e identificar las oportunidades.
- 4) **Observación directa versus observación indirecta.** La observación directa se refiere a la observación del comportamiento tal como ocurre realmente. La observación indirecta se refiere a la observación de algún registro del comportamiento pasado. En este caso se observan los efectos del comportamiento en vez de observar el comportamiento en sí. Esto abarca el examen de los rasgos físicos, un proceso que incluye aspectos tales como contar el número de envases de licor vacíos que aparecen en los depósitos de basura para estimar el consumo de licor en las familias.
- 5) **Observación humana versus observación mecánica.** En algunas situaciones es apropiado complementar o reemplazar al observador humano con algún tipo de observador mecánico.

La razón puede ser incrementar la precisión, disminuir los costos o requisitos especiales de medición. Entre los aparatos utilizados en la observación se incluye a la cámara de vídeo.

## **2. Método de comunicación**

El método de comunicación de recolección de datos se basa en la interrogación de los encuestados. Es lógico formular preguntas a los encuestados si se desea conocer cuál es la marca de jabón que compran, cuáles son los programas de televisión que ven o por qué compran en una determinada distribuidora. Tales preguntas pueden formularse verbalmente o por escrito y las respuestas pueden presentarse en cualquiera de las dos formas. El instrumento de recolección de datos que se utiliza en este proceso se llama cuestionario. El cuestionario se ha convertido en el instrumento predominante de recolección de datos en muchas investigaciones en distintas áreas de la empresa.

### **a. Ventajas del método de comunicación.**

La principal ventaja del método de comunicación es su versatilidad. La versatilidad se refiere a la capacidad del método para recolectar datos sobre una amplia gama de necesidades de información. La mayor parte de los problemas de decisión en mercadeo por ejemplo involucran personas. Por consiguiente, las necesidades de información se centran en el comportamiento pasado, actitudes y características de las personas. El método de comunicación puede recolectar datos en estas tres áreas.

Las ventajas adicionales se relacionan con la velocidad y el costo del método de comunicación, comparado con el método de observación. Las ventajas de velocidad y costo son altamente interdependientes. El método de comunicación es un medio más rápido de recolección de datos que el método de observación, puesto que proporciona un mayor control sobre el proceso de recolección de datos. El investigador no tiene que predecir cuándo y dónde va a ocurrir el comportamiento, ni esperar que ocurra. Por ejemplo, sería más rápido y más barato preguntar al encuestado acerca de la compra de un producto que tratar de anticipar y observar la compra.

### **b. Desventajas del método de comunicación.**

Existen varias limitaciones importantes del método de comunicación. La primera tiene que ver con la renuencia del encuestado a suministrar los datos deseados. Este puede negarse a ceder su tiempo para que lo entrevisten o para responder determinadas preguntas.

La segunda limitación se relaciona con la incapacidad del encuestado para suministrar los datos. Es posible que el encuestado no recuerde los hechos en cuestión o que ni siquiera sean de su conocimiento.

La limitación final está relacionada con la influencia del proceso de interrogación en las respuestas. Es posible que los encuestados sesguen sus respuestas con el propósito de dar una respuesta socialmente aceptable o para complacer al entrevistador.

Aunque las limitaciones pueden reducir seriamente la validez del método de comunicación, también pueden controlarse mediante un diseño adecuado del instrumento de recolección de datos.

A pesar de lo antes señalado, el método de comunicación es el medio más aplicado en las investigaciones empresariales, es por tal motivo que conviene estudiar algunas de las técnicas de comunicación con base a las clasificaciones por estructura y grado de ocultamiento del objetivo del estudio.

### c. Clasificación de las técnicas de comunicación

Existen varias formas de clasificar las técnicas de recolección de datos que emplean el método de comunicación. Un esquema de clasificación útil es categorizar las técnicas por su grado de estructura y el grado de ocultamiento del objetivo del estudio. Los cuatro tipos de técnicas que surgen en base a los dos criterios de clasificación son:

- 1) Estructurado - directo;
- 2) No estructurado - directo;
- 3) No estructurado - indirecto y
- 4) Estructurado - indirecto.

En la figura 7 se muestra un resumen de las técnicas de comunicación por tipo de estructura y grado de ocultamiento del objetivo de estudio.

**Técnicas de comunicación por tipo de estructura y grado de ocultamiento del objetivo de estudio**

Figura 7

	Estructurado	No estructurado
Directo	Cuestionario de la encuesta	Entrevista de grupos foco Entrevista en profundidad
Indirecto	Técnica del desempeño de la tarea objetiva	Técnicas proyectivas

Fuente: Thomas C. Kinnear y James R. Taylor, Investigación de Mercados. Un enfoque aplicado, 1994. Página 286.

La estructura se refiere al grado de normalización impuesto al proceso de recolección de datos. Generalmente se utiliza un cuestionario formal para darle estructura al proceso de recolección de datos. Con un enfoque altamente estructurado, las preguntas formuladas y las respuestas permitidas están completamente predeterminadas.

#### ***Veamos los siguientes ejemplos:***

- *Un examen de selección múltiple en la asignatura de técnicas de recopilación de la información es un ejemplo de un cuestionario altamente estructurado.*
- *Otro ejemplo de preguntas altamente estructuradas, lo constituyen:*  
*¿De qué Región proviene Usted? : Pacífico:\_\_\_ Centro:\_\_\_ Atlántico:\_\_\_*  
*¿En que tipo de colegio estudio usted?: Público:\_\_\_ Privado:\_\_\_*

*Observemos que las respuestas a estas preguntas ya están determinadas, no se obtendrán otras respuestas que las antes previstas.*

Con un enfoque altamente no estructurado de recolección de datos, las preguntas que se formulan se predeterminan vagamente y los encuestados tienen la oportunidad de expresar sus puntos de vista libremente.

**A manera de ejemplo:**

*Las preguntas siguientes son altamente no estructuradas:*

*¿Qué opina usted del aborto?: \_\_\_\_\_*

*¿Qué opinión le merece a usted la asignatura de Técnicas de Recopilación de la Información?: \_\_\_\_\_*

*Como se logra apreciar las respuestas no están previamente determinadas, existiendo un alto grado de decisión por parte de la persona para contestar lo que estime conveniente.*

**Me ejercito:**

*Elabore algunas preguntas que se me ocurran con formatos altamente estructurado y altamente no estructurado, lo comparto con mis compañeros.*

Un grado de estructura intermedia implicaría las preguntas normalizadas, pero con un formato de respuesta "abierta".

Un enfoque directo de recolección de datos es aquel en el cual el objetivo del estudio es obvio para el encuestado, de acuerdo con la naturaleza de las preguntas formuladas. En contraste, un enfoque indirecto formula preguntas de una manera que oculta el objetivo del estudio.

## **1) Estructurado - directo.**

La técnica más común de recolección de datos es el cuestionario estructurado - directo el cual exige que las preguntas se formulen para todos los encuestados con las mismas palabras exactamente y en la misma secuencia. Esta normalización se diseña para controlar el sesgo en las respuestas, asegurando que los encuestados estén respondiendo exactamente a la misma pregunta. La respuesta a una pregunta normalizada puede requerir la selección de una alternativa tomada de un conjunto predeterminado de respuestas.

Las preguntas normalizadas y las alternativas de respuesta fija pueden desarrollarse a partir de una investigación previa que se denomina piloto, en la cual se utilizaron técnicas menos estructuradas. Además, las categorías estructuradas de preguntas (pretest) y respuestas generalmente requieren extensas pre - pruebas para asegurar que las preguntas midan lo que deben medir.

El cuestionario estructurado-directo requiere de tiempo y grandes habilidades para poder desarrollarlo.

Son varias las ventajas del enfoque estructurado-directo, la más importante es su simplicidad para administrarlo y la facilidad con la cual pueden procesarse, analizarse e interpretarse los datos. El formato de preguntas estructuradas está diseñado para controlar sesgos en las respuestas e incrementar la confiabilidad de los datos.



Las desventajas del método estructurado-directo son las mismas que las limitaciones del método de comunicación: los encuestados pueden no estar en capacidad de proporcionar los datos deseados, es posible que no deseen suministrarlos y el proceso de interrogación puede sesgar sus respuestas. Asimismo, las preguntas estructuradas y las alternativas de respuestas fijas generarían una pérdida en la validez de algunos tipos de datos.

En el cuarto tema de la presente unidad autoformativa se profundiza en la elaboración del cuestionario.

## **2) No estructurado - directo.**

En el enfoque no estructurado- directo, el objetivo del estudio de investigación es claro para el encuestado. Hay un amplio grado de flexibilidad en la forma como se formulan las preguntas y en el grado de indagación. El formato de respuestas es abierto y los encuestados tienen la oportunidad de expresar libremente sus persuasiones y sentimientos sobre los temas presentados por el entrevistador. Hay dos técnicas que utilizan el enfoque no estructurado-directo:

- a). Entrevista de grupos foco y
- b). Entrevista en profundidad.

a). *Entrevista de grupos foco*. La entrevista de grupos foco y la entrevista de grupos en profundidad, son dos técnicas que se emplean con mucha frecuencia en las investigaciones sociales, aplicándose también en investigación de mercados.

La entrevista de grupo focal, puede definirse como una entrevista vagamente estructurada, dirigida por un moderador entrenado entre un número pequeño de encuestados simultáneamente. Esta técnica tiene sus orígenes en los métodos de la terapia de grupo utilizados por los psiquiatras. Aunque la entrevista requiere de una organización de los temas o guía del entrevistador, antes de elaborarla, el escenario hace énfasis en que la flexibilidad y el valor de la técnica está en descubrir lo inesperado, lo cual resulta de una libre discusión en grupo.

Las entrevistas de grupos foco pueden emplearse para diferentes objetivos, entre otros están:

- Generar hipótesis que pueden probarse más adelante en forma cuantitativa.
- Generar información útil en la estructuración de los cuestionarios del consumidor.
- Generar ideas para nuevos conceptos creativos.
- Interpretar resultados cuantitativos obtenidos previamente.

El grupo de encuestados debe estar compuesto por personas con características bastante homogéneas.

El tamaño del grupo puede ser de 10 ó 12 personas.

Con relación a la selección de los encuestados, la mayoría de los investigadores considera que es esencial realizar una selección cuidadosa para que la entrevista de grupos foco sea todo un éxito.

El papel del moderador es de gran importancia para el éxito de la técnica de grupo foco. Los moderadores altamente calificados pueden asegurar que se establezca una armonía adecuada con los encuestados, que la discusión se oriente hacia campos apropiados y que el grado de indagación y profundidad de comprensión sean suficientes para lograr los objetivos de la investigación.

Si se compara con otras técnicas de recolección de datos, la entrevista de grupo foco tiene las siguientes ventajas específicas:

- Sinergismo: el efecto combinado del grupo generará una gama más amplia de información, comprensión e ideas, que la que puede generar la acumulación de respuestas de varios individuos, cuando éstas se dan en forma privada.
- Crecimiento continuo: un efecto de adhesión a la multitud se presenta en una situación de entrevista de grupo, puesto que un comentario de un individuo a menudo desencadena una serie de respuestas de los otros participantes.
- Estimulación: con frecuencia, después de un breve periodo de introducción, los encuestados "entran en ambiente" puesto que desean expresar sus ideas y exponer sus sentimientos a medida que aumenta en el grupo el nivel general de entusiasmo sobre el tema.
- Seguridad: generalmente, los participantes se encuentran a gusto en el grupo cuando sus sentimientos no se diferencian demasiado de los de los otros integrantes y están más deseosos de expresar sus ideas y sentimientos.
- Espontaneidad: puesto que a los individuos no se les exige que respondan ninguna de las preguntas en la entrevista de grupo, sus respuestas pueden ser más espontáneas y menos convencionales, y deben proporcionar un cuadro más preciso de su posición con relación a algunos temas.
- Descubrimiento casual: se presenta con mayor frecuencia en una entrevista de grupo que en una entrevista individual, el hecho de que alguna idea "caiga del cielo".
- Especialización: la entrevista de grupo permite que se utilice un entrevistador altamente entrenado pero más costoso puesto que se está "entrevistando" a varios individuos simultáneamente.
- Escrutinio científico: la entrevista de grupo permite que se haga un escrutinio más detallado del proceso de recolección de datos, puesto que varios observadores pueden presenciar la sesión y puede grabarse para un análisis y reproducción posteriores.
- Estructura: la entrevista de grupo da una mayor flexibilidad que la entrevista individual, tanto en lo que se refiere a los temas cubiertos como a la profundidad con la cual se tratan los mismos.
- Velocidad: debido a que se entrevistan varios individuos simultáneamente, la entrevista de grupo acelera el proceso de recolección y análisis de datos.

La entrevista de grupos foco es una técnica de investigación exploratoria que puede ser muy valiosa para desarrollar hipótesis sobre problemas y oportunidades facilitando el desarrollo de una exposición clara del problema de decisión y estimulando el proceso creativo diseñado para formular cursos alternativos de acción.

- b). *Entrevista en profundidad.* La entrevista en profundidad puede definirse como una entrevista personal no estructurada que utiliza una indagación exhaustiva para lograr que un solo encuestado hable libremente y exprese en forma detallada sus persuasiones y sentimientos sobre un tema. El objetivo de esta técnica es llegar más allá de las reacciones superficiales del encuestado y descubrir las razones más fundamentales de sus actitudes y comportamiento.

La entrevista en profundidad puede extenderse por una hora o más con el entrevistador quien ha memorizado los temas que se han de tratar. Las palabras empleadas en las preguntas y su secuencia se dejan a discreción del entrevistador, quien trata de identificar las áreas generales de discusión para luego estimular al encuestado a que trate libremente, en profundidad, el tema de interés. El entrevistador indaga sobre las respuestas que son de interés formulando preguntas tales como: Eso es interesante, "¿podría hablarme más acerca de ello?" y "¿Por qué dice usted esto?"

Aunque no es un participante activo en la discusión, el entrevistador desempeña un papel crítico para el éxito de la técnica de la entrevista en profundidad. Es responsabilidad del entrevistador crear un ambiente en el cual el encuestado se sienta relajado y libre para presentar sus sentimientos y persuasiones sin miedo a la crítica o a la falta de comprensión.

Al igual que la entrevista de grupo foco, el papel del entrevistador es fundamental para el éxito de esta técnica.

La ventaja de la entrevista en profundidad sobre la entrevista de grupo foco radica en la mayor profundidad de persuasión que puede lograrse y habilidad para asociar la respuesta directamente con el encuestado. En la entrevista de grupo foco, es difícil determinar qué encuestado respondió de una determinada forma. Además, el entrevistador puede desarrollar un alto nivel de armonía con el encuestado, dando como resultado un intercambio más libre de respuestas, de lo que es posible con la técnica de grupo foco.

La técnica es útil para desarrollar hipótesis, definir problemas de decisión y formular cursos de acción.

El tiempo de duración de la entrevista, combinada con el costo del entrevistador, da como resultado un número pequeño de personas entrevistadas en el proyecto. El reducido tamaño de la muestra y la completa dependencia respecto del entrevistador para el análisis y la interpretación de la información, son limitaciones importantes que restringen el uso de esta técnica a situaciones de problemas especiales.

### **3) No estructurado-indirecto.**

Las técnicas no estructuradas-indirectas de recolección de datos reciben el nombre de técnicas proyectivas. Se derivan de la psicología clínica y se diseñaron para obtener datos indirectos sobre las persuasiones y sentimientos de los encuestados. Las técnicas estructuradas-directas de recolección de datos suponen que los encuestados comprenden claramente sus persuasiones y sentimientos, y desean comunicar estos datos en forma directa. Sin embargo, en algunas situaciones, este puede no ser el caso.

Las técnicas proyectivas se diseñaron para explorar los "por qué" del comportamiento. Una técnica proyectiva puede definirse como cualquier forma indirecta de formular una pregunta, en la que se crea un ambiente que estimula al encuestado a proyectar libremente persuasiones y sentimientos sobre el tema de interés.

Las técnicas proyectivas exigen que los encuestados interpreten el comportamiento de otros, en vez de solicitarles directamente que informen sus persuasiones y sentimientos. Al interpretar el comportamiento de otros, los encuestados están proyectando indirectamente sus propias persuasiones y sentimientos en la situación.

Aunque la mayor parte de las técnicas proyectivas se administran e interpretan fácilmente, su uso en la investigación se ha limitado a situaciones especiales en las cuales no pueden medirse directamente las actitudes.

Las técnicas proyectivas, como la entrevista en profundidad, presentan grandes desventajas que contribuyen a su uso limitado en la investigación en el área de administración. Debido a que a menudo requiere entrevistas personales con entrevistadores e intérpretes altamente capacitados en el análisis de datos, tiende a ser un instrumento de investigación muy costoso. El alto costo por cada encuestado que se entrevista ha dado como resultado el uso de tamaños de muestras pequeñas no probabilísticas, lo cual genera grandes errores muestrales. Las técnicas proyectivas son complejas y pueden conducir a conclusiones erróneas en manos de un principiante.

A pesar de estas limitaciones, se ha encontrado que las técnicas proyectivas son útiles en muchas situaciones. Su uso adecuado todavía radica en la investigación exploratoria diseñada para descubrir hipótesis que puedan probarse mediante técnicas de investigación más estructuradas y directas.

### **4) Estructurado-indirecto.**

Con frecuencia, el enfoque estructurado-indirecto se conoce como técnica del desempeño de la tarea objetiva. Se solicita a los encuestados que memoricen y/o presenten la información basada en hechos acerca del tema de interés. Se analizan estas respuestas y se hacen inferencias acerca de la naturaleza de las persuasiones y sentimientos fundamentales de los encuestados con relación al tema.

El supuesto central de esta técnica es que los encuestados recuerdan con mayor facilidad aquellas cosas que son afines con sus persuasiones y sentimientos. Presumiblemente, un mayor conocimiento refleja la fuerza y dirección de los otros componentes de la actitud. Este argumento se basa en los hallazgos de la investigación sobre el procesamiento de la

información selectiva, hallazgos que indican que las personas tienden a: Exponerse selectivamente a la información, percibir selectivamente la información y retener selectivamente la información que es compatible con sus actitudes. Por consiguiente, solicitar a los encuestados que recuerden la información basada en hechos acerca de un tema, es una forma indirecta de medir la orientación y fuerza de sus actitudes. Se supone que aquellos encuestados que tengan un mayor conocimiento acerca del producto X tienen actitudes más fuertes hacia éste.

El enfoque estructurado-indirecto rara vez se utiliza en la investigación en administración.

Este representa un intento para tratar de obtener las ventajas de la medición indirecta de actitudes con las ventajas del procesamiento y recolección de datos de los enfoques estructurados. Muchos investigadores se preocupan acerca de si la medición de información basada en hechos es un indicador válido de la dirección y fuerza de la estructura implícita de la actitud.

### 3. Medios para recolectar datos por el método de comunicación

Los tipos de medios de comunicación disponibles para obtener datos de los encuestados se muestran en la figura 8:

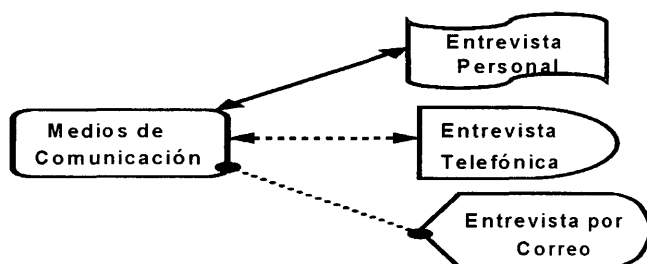


Figura 8

Siendo estos:

- a. Entrevista personal,
- b. Entrevista telefónica y
- c. Entrevista por correo.

Estos tres medios de comunicación son ampliamente usados en la investigación.

Las técnicas de comunicación estructuradas pueden emplear estos tres medios. Las técnicas no estructuradas por lo general requieren de la entrevista personal. La entrevista telefónica se puede utilizar con éxito en la técnica de grupo foco, por medio de teleconferencia.

Con frecuencia, las entrevistas telefónicas y por correo no se prestan a buenos resultados cuando utilizan las técnicas no estructuradas.

### **a. Entrevista personal.**

La entrevista personal consiste en que un entrevistador formula preguntas a uno o más encuestados en una situación cara a cara. La tarea del entrevistador es ponerse en contacto con el encuestado, formular las preguntas y registrar las respuestas. Las preguntas deben formularse en forma clara y registrarse con mucha exactitud. El registro de las respuestas puede hacerse durante o después de la entrevista. El proceso de entrevista cara a cara puede hacer que los encuestados sesguen sus respuestas por ejemplo por el deseo de complacer o impresionar al entrevistador.

*Un ejemplo concreto de entrevista personal para obtener datos de los encuestados fue realizado en el último censo nacional realizado en nuestro país en 1995 por el INEC(Instituto Nacional de Estadísticas y Censos).*

### **b. Entrevista telefónica.**

La entrevista telefónica consiste en que un entrevistador formula preguntas a uno o más encuestados a través del teléfono en vez de un contacto personal directo. La entrevista telefónica es la que más se utiliza de los tres medios de comunicación en los países desarrollados. Su popularidad se debe tanto a sus procedimientos eficientes como económicos y a su aplicación en una amplia gama de necesidades de información. Esto no ocurre en nuestro país, principalmente por el poco número de líneas telefónicas existentes y por una concentración de estas en las ciudades con mayor concentración poblacional.

El menor grado de interacción social que se presenta entre el entrevistador y el encuestado en la entrevista telefónica, reduce el potencial de sesgo en comparación con la entrevista personal. Las limitaciones básicas de la entrevista telefónica se relacionan con la cantidad limitada de datos que puede obtenerse y al sesgo potencial que puede resultar de una lista incompleta de la población objetivo, es decir, teléfonos que no aparecen en la lista y que no pertenecen a los propietarios del inmueble entre otras.

*Un ejemplo concreto de entrevista telefónica para obtener datos de los encuestados podrá ser utilizado por Bell South,, interesada la dirección de la empresa en conocer la percepción de sus clientes en cuanto al servicio que brinda.*

### **c. Entrevista por correo.**

La entrevista por correo consiste en un cuestionario enviado por correo al encuestado y la devolución por correo del cuestionario completo a la organización de la investigación. Es tan popular como la entrevista personal, pero menos que la entrevista telefónica.

Las entrevistas por correo son flexibles en su aplicación, relativamente bajas en su costo y carecen del potencial de sesgo como resultado de la interacción entrevistador-encuestado.

La mayor desventaja se relaciona con el problema del error de no respuesta. Para distribuir y recolectar el cuestionario pueden utilizarse muchos enfoques, una persona

puede encargarse de entregarlos y/o recogerlos en lugar de ser enviados por correo, puede distribuirse en revistas y periódicos, pueden adjuntarse tarjetas de garantía a los productos y de esta manera recolectar los datos sobre las características del comprador y del proceso de decisión de compra.

*Un ejemplo concreto de entrevista por correo para obtener datos de los encuestados, es realizado por la firma de auditores Price Waterhouse Coopers entre diversos clientes de Banco Uno para realizar una auditoría en los cuenta-habientes de dicho banco.*

**Me ejercito:**

*De acuerdo a mi experiencia ejemplifico los tres tipos de entrevistas estudiados para recolectar datos por el método de comunicación. Me preparo para socializarlo con el resto de los compañeros.*

#### **4. Criterios para seleccionar los medios de comunicación**

Son varios los criterios pertinentes en la evaluación de los mejores medios de comunicación que satisfacen las necesidades de un proyecto de investigación, específicamente: versatilidad, costo, tiempo, control de la muestra, cantidad de datos, calidad de los datos y tasa de respuesta. La importancia asignada a cada uno de estos criterios varía de acuerdo con las necesidades específicas del proyecto de investigación.

- **Versatilidad:** La versatilidad se refiere a la habilidad del medio para adaptar el proceso de recolección de datos a las necesidades especiales del estudio o del encuestado. La entrevista personal es el más versátil de los tres medios de comunicación. Las entrevistas telefónicas son menos versátiles que las entrevistas personales, mientras que la entrevista por correo es la menos versátil.

El investigador debe determinar el grado de versatilidad requerido en un proyecto de investigación y seleccionar los medios de comunicación que mejor satisfagan las necesidades del estudio. En la práctica la mayor parte de los proyectos de investigación no requieren de la alta versatilidad que ofrece la entrevista personal.

- **Costo:** El número de horas de trabajo tiende a determinar el costo relativo de los tres medios de comunicación. Los costos de la mano de obra incluyen los salarios de los entrevistadores y los costos de supervisión asociados al control de calidad del proceso de recolección de datos.

Generalmente la entrevista personal es el medio más costoso por entrevista terminada. Por lo regular las entrevistas telefónicas son más costosas que las entrevistas por correo. Cuando el cuestionario es breve, el costo de la entrevista telefónica usualmente es igual al de la entrevista por correo.

- **Tiempo:** De los tres medios de comunicación, la entrevista telefónica es la forma más rápida para obtener datos. Con un cuestionario breve, un entrevistador puede completar entre 10 o más entrevistas por hora. Utilizando el mismo cuestionario, un entrevistador personal podría considerarse afortunado al completar dos o tres entrevistas por hora. Sin duda el tiempo de viaje entre las entrevistas representa una seria restricción de tiempo en la tasa de terminación del entrevistador personal. Por

consiguiente, los estudios basados en entrevistas personales, por lo general son más prolongados en el tiempo que transcurre entre el inicio del trabajo de campo y la finalización del proyecto, en comparación con los estudios basados en entrevistas telefónicas o por correo.

- **Control de la muestra:** El control de la muestra se refiere a la capacidad de los medios de comunicación para alcanzar las unidades designadas en el plan de muestreo, efectiva y eficientemente. Los tres medios de comunicación se diferencian en forma significativa a este respecto. La entrevista personal ofrece el mejor grado de control de la muestra.

El marco muestral es la lista de las unidades poblacionales de la cual se tomará la muestra. Los procedimientos muestrales que no requieren de una lista de unidades se basan, en gran parte, en el entrevistador personal durante el proceso de selección de la muestra. Trabajando a través del entrevistador personal el investigador puede controlar las unidades muestrales que se han entrevistado, las personas entrevistadas, el grado de participación de otros miembros de la unidad en la entrevista y muchos otros aspectos del proceso de recolección de datos.

La entrevista telefónica depende fundamentalmente de un marco muestral. Usualmente, uno o más directorios telefónicos sirven como marco muestral con encuestados seleccionados de los directorios que representan la población de interés, utilizando procedimientos probabilísticos de selección.

Con frecuencia los directorios telefónicos representan marcos muestrales muy pobres, puesto que son listas incompletas de personas en un área.

Debido a que los directorios telefónicos no son representativos, la entrevista telefónica sólo permite un control limitado sobre la muestra.

La entrevista por correo, así como la entrevista telefónica, requieren de una lista de los elementos poblacionales. Idealmente este marco está compuesto por nombres y direcciones.

- **Cantidad de datos:** Una regla establecida es que se puede recolectar la mayor cantidad de datos utilizando la entrevista personal seguida por la entrevista por correo y luego la entrevista telefónica. Existe creciente evidencia de que la entrevista por correo y la telefónica pueden recolectar más datos de lo que se suponía. En situaciones en las cuales los encuestados están emocionalmente involucrados en el tema, los tres medios de comunicación pueden proporcionar cantidades sustanciales de datos. Sin embargo, a niveles normales de implicación la entrevista personal puede recolectar sustancialmente más datos que las otras dos.

La principal ventaja de la entrevista personal radica en la relación social entre el entrevistador y el encuestado. Generalmente este ambiente social motiva al encuestado a dedicar más tiempo a la entrevista. Para el encuestado es más fácil finalizar la entrevista telefónica o por correo debido a su naturaleza impersonal, que finalizar la entrevista personal.



Una ventaja de las entrevistas personales y telefónicas con relación a la entrevista por correo, es que se requiere de un menor esfuerzo del encuestado en el proceso de recolección de datos. En este caso el entrevistador formula las preguntas, profundiza las respuestas y registra los resultados. La entrevista personal tiene la ventaja adicional de permitir la presentación visual de las categorías de escalas de clasificación y de otro material de apoyo que le puede facilitar al encuestado la comprensión de las preguntas formuladas. Todas estas ventajas contribuyen a una mejor disposición del encuestado para suministrar mayor cantidad de datos.

- **Calidad de los datos:** La calidad de los datos se refiere al grado en el cual éstos están libres de un sesgo potencial como resultado del uso de un determinado medio de comunicación. Por lo general cuando el tema no es emocional y el cuestionario se diseña y administra adecuadamente los datos serán de calidad, independientemente del medio que se emplee.

Los investigadores han encontrado diferencias significativas entre los tres medios cuando utilizan preguntas sensibles o embarazosas; por ejemplo, préstamos bancarios, ingresos, o comportamiento sexual. En el caso de temas sensibles, existe evidencia de que las encuestas por correo recolectan datos de mejor calidad que las entrevistas personales. Las entrevistas telefónicas se situarían en un punto intermedio entre estos dos medios.

Otra fuente de sesgo resulta de la confusión de los encuestados con la pregunta formulada. En una entrevista por correo el encuestado no puede hacer que el entrevistador le clarifique la pregunta; este tipo de entrevista ofrece la mayor oportunidad para que los resultados sean inexactos debido a la confusión. La entrevista telefónica ofrece un mayor potencial de sesgo que la entrevista personal, debido a que no hay presencia física del entrevistador.

La entrevista por correo tiene otro sesgo potencial que puede resultar debido a que el encuestado lee todas las preguntas antes de contestarlas o cambia las respuestas al comienzo del cuestionario como resultado de respuestas dadas posteriormente. Tanto la entrevista personal como la telefónica están expuestas a este tipo de sesgo.

La calidad de los datos obtenidos en las entrevistas telefónicas tiende a ser mejor que los de las entrevistas personales, puesto que el proceso de recolección de datos puede supervisarse y controlarse más eficientemente. Se necesitan menos entrevistadores para la encuesta telefónica y pueden capacitarse y supervisarse centralmente.

Tanto la entrevista telefónica como la personal tienen una importante ventaja sobre la entrevista por correo puesto que pueden utilizarse para recolectar datos en el momento en que ocurre el comportamiento. Esto reduce el sesgo asociado con el fracaso de recordar eventos con precisión.

Una consideración final con respecto a la calidad de los datos obtenidos está relacionada con las trampas que pueda hacer el entrevistador. Es muy fácil hacer trampa en la entrevista personal, puesto que la entrevista telefónica puede monitorearse directamente, mientras que la primera no. Debido a que la entrevista por correo no tiene un entrevistador, no está sujeta al sesgo que resulta del engaño.

- **Tasa de respuesta:** La tasa de respuesta se refiere al porcentaje de la muestra original que en realidad se entrevista. Una baja tasa de respuesta puede resultar en un alto error de no respuesta, lo cual puede invalidar los hallazgos de la investigación. El error de no respuesta se refiere a la diferencia entre los que responden a una encuesta y aquellos que no responden.

El error de no respuesta es una de las fuentes más graves de error que tiene que enfrentar el investigador.

La probabilidad del error de no respuesta aumenta cuanto menor sea la tasa de respuesta de la encuesta.

Sin embargo, es importante reconocer que una tasa de respuesta baja en sí no implica que no exista un alto error de no respuesta en los datos, sólo cuando existe una diferencia en las variables de interés entre los encuestados y los no encuestados se presenta un error de no respuesta. Si la razón de la no respuesta es independiente de las variables claves de interés, debe ser muy poca la diferencia entre los grupos de encuestados y de no encuestados.

La no respuesta puede resultar de dos fuentes: No se encuentra en casa y rechazos. La no respuesta debido a que el encuestado no se encuentra en casa puede afectar seriamente las entrevistas telefónicas y las personales, pero tiene una influencia limitada en las entrevistas por correo.

- **Selección del medio de comunicación.** ¿Qué medio de comunicación debe seleccionarse para un estudio? La respuesta está en el medio más adecuado que satisfaga las necesidades de información del estudio dadas las limitaciones de tiempo y costos. Debe destacarse que los tres medios de comunicación no son mutuamente excluyentes. Pueden utilizarse varias combinaciones de medios en el diseño de investigación; esto permite que el investigador diseñe un estudio que aproveche las ventajas de los tres medios.

A manera de síntesis, en este apartado hemos estudiado el cómo podemos recolectar los datos de interés, identificando claramente la existencia de dos formas o métodos de realizarlo que son: La Observación y La Comunicación.

En el caso de la observación, este método se asocia a lo que observamos del mundo que nos rodea, es decir lo que vemos en cuanto a comportamiento o fenómenos objeto de nuestro estudio. Por otro lado en el caso del método de Comunicación, necesariamente debemos de realizar la interrogación a las personas de lo que queremos investigar.

Hemos establecido diversas técnicas para cada uno de los métodos estudiados. Aunque es conveniente tener presente que en dependencia del tipo de investigación que realicemos, podremos realizar una combinación de técnicas de los dos métodos estudiados.

Un método que le damos mucha importancia en el marco de lo que estamos estudiando en el presente módulo autoformativo, lo constituye el método de la comunicación con la técnica del cuestionario (encuesta), la cual se ubica en la clasificación de las técnicas estructurada – directa. Lo anterior se sustenta en la necesidad del trabajo de fin de curso que se orienta en la presente unidad autoformativa.

Para la técnica del cuestionario (encuesta) se ha realizado una revisión de las distintas formas o tipos de medios de comunicación que podemos utilizar para obtener los datos de interés, hemos estudiado la entrevista personal donde se destaca la existencia de un encuestador de forma directa con el entrevistado, la entrevista telefónica y la entrevista por correo. Todos estos medios podemos disponerlos para realizar la recopilación de los datos, existiendo la posibilidad de que realicemos combinaciones de ellos a partir de lo que planeemos en el estudio.

Se han estudiado los criterios que podemos tener presente para la elección de uno u otro medio.

Continuando con lo previsto en la presente unidad autoformativa, continuaremos el estudio de las encuestas por muestreo, más concretamente realizaremos una descripción de los pasos que debemos tener presente para la planeación de una encuesta por muestreo. Debemos tener presente algunos de ellos dado que en base a estos pasos prepararemos el plan del trabajo final, otros nos servirán para la ejecución del plan.

Como último aspecto estudiaremos lo relacionado al diseño de los cuestionarios que es el instrumento que tendremos que utilizar para que de cada elemento de la población elegido en la muestra le recopilemos los datos que nos servirán para el estudio que realizamos.

### **Actividad de Autoaprendizaje No. 3**

1. Establezco las ventajas y desventajas que se aprecian en el método de observación.
2. Expreso mi opinión sobre la entrevista a profundidad, caracterizándola previamente.
3. ¿Qué habilidades debe poseer quien entrevista?
4. Señalo las ventajas y desventajas que posee el cuestionario (encuesta) como parte de las técnicas estructurada- directa que hemos estudiado.

Luego comparo mis respuestas con las que aparecen en las páginas No. 116 y 117, al final de la unidad autoformativa I.

## D. ENCUESTAS POR MUESTREO

### 1. Planeación de una encuesta por muestreo<sup>2</sup>

Al planear una encuesta es conveniente realizar los pasos en el orden secuencial en que los mostramos en el presente subtema con el objetivo de asegurar todos los elementos necesarios para tener un adecuado diseño de lo que estemos interesados en estudiar. Cada aspecto por lo tanto, debe ser cuidadosamente considerado en la planeación de cualquier encuesta, aunque en el avance de dicha preparación, elementos que ya se han desarrollado podrán ser mejorados una vez que se profundice en el proceso de planeación.

La figura 9 nos muestra en orden secuencial, los once pasos a considerar para la planeación de una encuesta por muestreo.

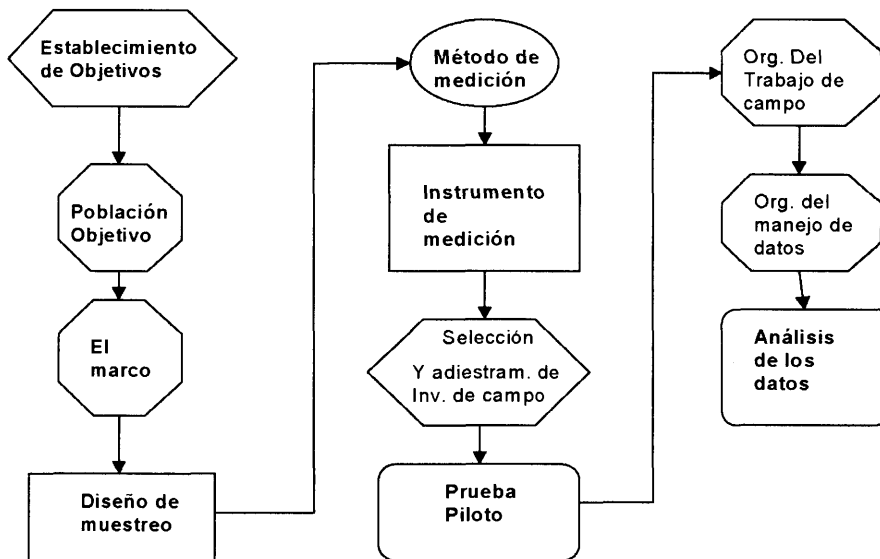


Figura 9

Atenderemos cada uno de los pasos para la planeación de una encuesta por muestreo:

#### a. Establecimiento de objetivos.

Establezca los objetivos de la encuesta de una manera clara y concisa, y remítase a estos objetivos conforme se vaya progresando en el diseño e instrumentación de la encuesta. Mantenga los objetivos suficientemente simples, de tal manera que sean entendidos por quienes trabajan en la encuesta y sean logrados exitosamente cuando se complete la encuesta.

<sup>2</sup> Parte del presente subtema ha sido tomado de Richard L. Scheaffer, William Mendenhall y Lyman Ott, 1987. Elementos de Muestreo. Tercera edición. Editorial Grupo Editorial Iberoamérica. México. Página 35 y 36.

Los objetivos responden a lo que estamos interesados en estudiar, podemos estructurar un objetivo general y de él desprender varios objetivos específicos.

Los objetivos deben ser redactados con un verbo en infinitivo que nos movilice a lo que queremos estudiar, adicionando lo que se estudiará.

***A manera de ejemplo se presenta lo siguiente:***

*Objetivo general: Conocer el nivel de incidencia de los hábitos de estudio en el rendimiento académico de los estudiantes de primer año de administración de empresas de la UCA en el año xxxx.*

*Observar que el verbo en infinitivo para este ejemplo nos motiva al conocimiento de una situación determinada, que en este caso es el nivel de incidencia de los hábitos de estudio en el rendimiento académico. Se le adiciona la población objeto de estudio, que en el objetivo general se asocia a "los estudiantes de primer año de administración de empresas de la UCA", así como también se le incluye el año o momento en el que la población se está estudiando.*

*Los objetivos generales entonces deben contemplar estos cuatro elementos en su redacción.*

*Algunos de los objetivos específicos que se pueden desprender son:*

*Identificar las formas en que estudian los estudiantes de primer año de administración de empresas.*

*Conocer las horas que le dedican al estudio los estudiantes de primer año de administración de empresas.*

*Diferenciar en función del sexo las formas de estudio de los estudiantes de administración de empresas.*

Evidentemente se podrán generar muchos más objetivos específicos de los aquí señalados. Es importante notar que del objetivo general estamos desprendiendo los objetivos específicos, estos últimos deben estar en correspondencia del objetivo general. Haciendo una analogía con un edificio de dos pisos, para subir al segundo piso se requiere de las escaleras, cada peldaño corresponderá a un objetivo específico, el subir al segundo piso corresponde al objetivo general.

***Me ejercito:***

*Elaboro un sistema de objetivos sobre un tema de mi interés.*

**b. Población objetivo.**

Defina cuidadosamente la población que va a ser muestreada. Si los adultos van a ser muestreados, entonces defina qué se entiende por adulto (todas aquellas personas con más de 18 años de edad, por ejemplo), y establezca qué grupo de adultos está incluido (todos los residentes permanentes de una ciudad, por ejemplo). Tenga presente que se va a obtener una muestra de esa población, y defina la población de tal manera que la selección sea factible.

**Ejemplo:**

*En el caso del ejemplo que hemos planteado relacionado a la incidencia de los hábitos de estudio, la población objetivo que de forma precisa hemos incluido en el objetivo general, se asocia a los estudiantes de primer año que estudian la carrera de Administración de Empresas en el año xxxx en la UCA.*

Observar que debemos detallar muy bien quienes son los elementos que cumplen la característica de interés para que al sacar la muestra y al realizar estimaciones quede definido claramente los elementos que pertenecen a la población que estamos estudiando. Ningún elemento que no pertenece a la población podrá incluirse en lo que nosotros estamos definiendo.

**c. El marco.**

Seleccione el marco (o marcos) de tal manera que la lista de unidades muestrales y la población objetiva concuerden lo más posible. Tenga en cuenta qué marcos múltiples pueden hacer el muestreo más eficiente. Por ejemplo, los residentes de una ciudad pueden ser muestreados de una lista de manzanas de la ciudad, relacionada con una lista de residentes dentro de las manzanas.

**Ejemplo:**

*Continuando con nuestro ejemplo de los hábitos de estudio, debemos de detallar las listas de todos los estudiantes de primer año que estudian la carrera de Administración de Empresas en el año xxxx en la UCA. Observar que el marco se refiere entonces a esa lista de estudiantes que cumplen con la condiciones de ser estudiantes de la UCA, estudian administración de empresas cursan el primer año en el año calendario xxxx. Por cualquier razón nos podemos encontrar que algunos elementos que son de la población no se nos listó en el marco, es en este sentido que la población objetivo es el conjunto de referencia y de él desprendemos el marco, es decir otro conjunto de elementos que listamos y que es un subconjunto del conjunto de referencia denominado población objetivo.*

El marco entonces lo podemos ver como la concreción de los elementos que son de la población, que se nos escapan algunos, pero que los que hemos podido listar nos servirán para que de él se seleccione la muestra.

**d. Diseño de muestreo.**

Seleccione el diseño de muestreo, incluyendo el numero de elementos en la muestra, de tal manera que la muestra proporcione suficiente información para los objetivos de la encuesta. Muchas encuestas producen poca o inútil información porque no fueron propiamente diseñadas. El diseño de muestreo será estudiado en la segunda unidad autoformativa y los tipos de diseño probabilístico en la segunda y tercera unidad autoformativa.

Deberemos entonces en este paso elegir el método de muestreo, como veremos en las próximas unidades autoformativas podremos elegir entre muestreo aleatorio simple, sistemático, estratificado y por conglomerados. De igual manera deberemos establecer el número de elementos que conformarán la muestra.

El diseño de muestreo se asocia al método o procedimiento de muestreo que aplicaremos para elegir los miembros de la población o el marco y que serán los que pasarán a formar parte de la muestra. Para ello deberemos elegir el tamaño de la muestra en correspondencia del procedimiento de muestreo que hemos seleccionado.

Existe para cada procedimiento una fórmula para encontrar el tamaño de (n) y que depende de tres factores:

- ♥ El grado de confianza que es establecido por el investigador
- ♥ El máximo error permisible que también es elegido por el investigador
- ♥ La variabilidad de la población, que debe ser medida por medio de la aplicación de una prueba (muestra) piloto (paso 8) y sustituir la varianza que se obtenga en la prueba piloto por la varianza poblacional en las fórmulas de cálculo del tamaño de la muestra.

Estos aspectos son explicados con detalle en el tema A de la segunda unidad autoformativa.

#### **e. Método de medición.**

Decida sobre el método de medición, usualmente uno o más de los siguientes métodos: entrevistas personales, entrevistas por teléfono, cuestionarios enviados por correo u observación directa. Observar que el método de medición ha sido estudiado en el tema C de la presente unidad autoformativa: Métodos de recopilación de datos.

#### **Ejemplo:**

*En el caso de los hábitos de estudio que hemos venido asumiendo, el método de medición más adecuado corresponde a la entrevista personal, deberemos de tener presente los elementos que conforman la muestra y a ellos estaremos contactando personalmente una vez que ejecutemos el plan.*

#### **f. Instrumento de medición.**

En conjunción con el paso anterior, especifique cuidadosamente qué mediciones van a ser obtenidas. Si se va a usar un cuestionario, planea las preguntas de tal manera que se minimice la no respuesta y el sesgo por respuesta incorrecta. El instrumento será estudiado en el presente punto con el título: Diseño de formas para la recolección de datos.

Las preguntas deberán de corresponder a lo que los objetivos persiguen, es así que para cada objetivo específico que nos propongamos, deberá existir una pregunta por lo menos que formará parte de lo que denominaremos cuestionario o instrumento de medición para una encuesta.

**Por ejemplo:**

*Si tenemos como objetivo específico: Diferenciar en función del sexo las formas de estudio de los estudiantes de primer año de Administración de Empresas. Las preguntas que podemos formularnos para el instrumento de medición podrán ser:*

*¿Sexo?: M\_\_\_ F\_\_\_*

*¿De que forma estudia usted?: Individual\_\_\_ En grupo:\_\_\_*

*¿Cuántas horas le dedica usted al estudio individual?\_\_\_*

*¿Y al estudio en grupo?\_\_\_*

*¿qué otra forma utiliza para estudiar?\_\_\_*

Nos podremos auxiliar con la matriz sugerida para la elaboración del plan y que aparece en las orientaciones para la elaboración del trabajo final, en la parte final de la presente unidad autoformativa.

**g. Selección y adiestramientos de investigadores de campo.**

Seleccione cuidadosamente y adiestre a los investigadores de campo, Después de que el plan de muestreo esté clara y completamente establecido, alguien debe recolectar los datos. Aquellos que van a reunir los datos, los investigadores de campo, deben ser cuidadosamente adiestrados sobre qué mediciones hacer y cómo hacerlas. El adiestramiento es especialmente importante si se usan entrevistas personales o por teléfono, porque la tasa de respuesta y la exactitud de las respuestas son afectadas por el estilo personal y el tono de voz del entrevistador.

**h. Prueba piloto.**

Seleccione una pequeña muestra para una prueba piloto. La prueba piloto es crucial, ya que permite a usted probar en el campo el cuestionario y otros instrumentos de medición, calificar a los entrevistadores y verificar el manejo de las operaciones de campo. Los resultados de la prueba piloto usualmente sugieren algunas modificaciones antes de realizar el muestreo a escala completa.

De igual forma, la prueba piloto servirá para que en función de la variable de mayor interés en el estudio, se realice la medición correspondiente y los cálculos de la varianza de esa muestra piloto, reemplacen la varianza poblacional y así poder estimar el tamaño de la muestra.

**i. Organización del trabajo de campo.**

Planee en detalle el trabajo de campo. Cualquier encuesta de gran escala involucra un gran número de personas trabajando como entrevistadores, coordinadores o personal dedicado al manejo de datos. Antes de que empiece la encuesta deben organizarse cuidadosamente los diferentes trabajos y establecerse claramente las líneas de autoridad.



## **j. Organización del manejo de datos.**

Elabore un esquema de cómo se debe manejar cada pieza de información para todas las etapas de la encuesta. Las grandes encuestas generan gran cantidad de información. Es por eso que un plan bien preparado para el manejo de los datos es de la mayor importancia. Este plan debe incluir los pasos para el proceso de los datos, desde el momento en que se hace una medición en el campo hasta que el análisis final ha sido completado. Se debe incluir también un esquema de control de calidad para verificar la correlación entre los datos procesados y los datos recolectados en el campo.

## **k. Análisis de los datos.**

Defina los análisis que deben realizarse. Estrechamente relacionado al paso anterior, este paso involucra la especificación detallada de los análisis que deben ser ejecutados. Se pueden también listar los temas que serán incluidos en el reporte final. Si usted considera el reporte final antes de que la encuesta sea llevada a cabo, puede tener más cuidado en la selección de los aspectos a ser medidos en la encuesta.

Si se siguen estos pasos diligentemente, la encuesta está lista para un buen comienzo y debe proporcionar información útil para el investigador.

En resumen, el objetivo de una encuesta por muestreo es hacer inferencias acerca de la población de interés, con base en la información contenida en una muestra. La población consiste en el conjunto de datos sobre los cuales deseamos hacer una inferencia, y está compuesta de elementos o piezas de información. Las unidades de muestreo son colecciones de elementos de la población que no presentan traslapes. El marco es una lista de unidades de muestreo que usamos para representar la población. La muestra es una colección de unidades de muestreo tomadas a partir del marco. Mediante los datos de la muestra, vamos a estimar ciertos parámetros de la población y fijar límites sobre nuestro error de estimación. La cantidad de información obtenida de la muestra puede ser controlada por el número de unidades muestrales extraídas de la población y el diseño de muestreo o método de recolección de datos usado.

Después que el diseño ha sido seleccionado, existen varios métodos para recolectar los datos de la muestra. Entrevistas personales, entrevistas por teléfono, observación directa y cuestionarios enviados por correo fueron métodos discutidos y evaluados como medios de reunir los datos en la muestra. Cada método tiene sus ventajas y limitaciones.

En el siguiente subtema discutimos la construcción real de los cuestionarios. De nuevo, enfatizamos la importancia de obtener información en la muestra que sea representativa de la población. Este problema es de importancia primordial cuando consideramos métodos de recolección de información.

## 2. Diseño de formas (cuestionarios) para la recolección de datos de los encuestados<sup>3</sup>

Un cuestionario es un plan formalizado, estructurado-directo para recolectar datos de los encuestados. La función del cuestionario es la medición. Los cuestionarios pueden utilizarse para medir: El comportamiento pasado, las actitudes y las características del encuestado.

Los aspectos involucrados en el diseño del cuestionario se relacionan directamente con el desarrollo y administración de las técnicas de medición de actitudes.

El error de medición es un problema intrínseco en la construcción del cuestionario. Cuando se formula una pregunta de preferencia sin proponer alternativas realistas, los resultados pueden carecer de sentido.

Esta diferencia significativa hace énfasis en la importancia de la redacción del cuestionario. El investigador debe preguntarse: "¿Miden las preguntas lo que se supone que deben medir?" (recordar que esto se asocia con el concepto de validez) Si la respuesta es negativa, existe un error de medición. Es importante controlar el error de medición en la elaboración de cuestionarios. Por lo tanto se necesita habilidad para diseñar el cuestionario de forma tal que las preguntas formuladas midan lo que supone deben medir.

### a. Los componentes del cuestionario

La figura 10 nos muestra los componentes del cuestionario.

Sección 1	Datos de Identificación
Sección 2	Solicitud de Cooperación
Sección 3	Instrucciones
Sección 4	Información solicitada
Sección 5	Datos de Clasificación

Figura 10

<sup>3</sup> Parte del presente tema ha sido tomado de Thomas C. Kinnear y James R. Taylor. 1994. Investigación de Mercados. Un enfoque aplicado. Cuarta edición. Editorial Mc Graw Hill. Colombia. Capítulo XI, página 309-325.

Generalmente un cuestionario tiene cinco secciones:

- 1) Datos de identificación
- 2) Solicitud de cooperación
- 3) Instrucciones
- 4) Información solicitada y
- 5) Datos de clasificación

- 1) **Los datos de identificación** generalmente ocupan la primera sección de un cuestionario y se relacionan con el nombre, dirección y número telefónico del encuestado. Los datos adicionales incluirían elementos tales como la hora y la fecha de la entrevista, además del nombre o número del código del entrevistador.
- 2) **La solicitud de cooperación** es una pequeña enunciación abierta diseñada para obtener la cooperación del encuestado con relación a la entrevista. Usualmente esta enunciación identifica primero al entrevistador y/o a la organización que está elaborando la entrevista. En seguida se explica el objetivo del estudio y se indica el tiempo que se requiere para completar la entrevista.
- 3) **Las instrucciones** se refieren a comentarios realizados al entrevistador o encuestado con relación a la forma de utilizar el cuestionario. Estos comentarios aparecen directamente en el cuestionario cuando se emplea una encuesta por correo. En el caso de la encuesta personal y telefónica se incluye una hoja por separado titulada: "Instrucciones del entrevistador" en la cual se explica el objetivo del estudio, plan de muestreo y otros aspectos del proceso de recolección de datos. Además, el cuestionario puede contener instrucciones especiales relacionadas con el uso de preguntas específicas, por ejemplo, la técnica de escala de actitudes.
- 4) **La información solicitada** constituye la parte más grande del cuestionario. En ella se incluyen todas las preguntas que deben ser contestadas por el entrevistado y que representan la esencia del cuestionario.
- 5) **Los datos de clasificación** tratan sobre las características del encuestado. Estos datos los suministra directamente el encuestado en el caso de una encuesta por correo. En las encuestas personales y las telefónicas el entrevistador recolecta los datos o, en algunos casos, puede estimar tipos más sensibles de datos basado en la observación, como por ejemplo, el ingreso.

A menudo los datos de clasificación se recolectan al final de la entrevista. Sin embargo, algunos procedimientos muestrales requieren que los datos de clasificación se recolecten al comienzo de la entrevista para determinar si la persona califica como parte del plan de muestreo.

### **Me ejercito:**

*Preparo un cuestionario para recolectar datos entre los trabajadores de la empresa donde laboro, relacionado a la satisfacción del ambiente laboral. Me preparo para socializarlo en el grupo de clase.*

### A manera de ejemplo:

Se presenta el siguiente cuestionario que trata de mostrar los componentes del mismo expresado en la figura 10:

Cuestionario Satisfacción de clientes para una empresa de Servicios

Sexo: H\_\_\_\_ M\_\_\_\_

Edad: \_\_\_\_\_

Para servirle mejor, nos gustaría conocer su opinión respecto a la calidad de nuestros servicios. Favor indique hasta que punto está de acuerdo o en desacuerdo con las siguientes aseveraciones sobre el servicio que recibió por parte de nuestra empresa. Favor encierre en un círculo el número que corresponda y utilizando la escala que se indica a continuación: 1: Totalmente en desacuerdo, 2: en desacuerdo, 3: ni de acuerdo/ ni en desacuerdo, 4: de acuerdo, 5: Totalmente de acuerdo.

	TD	D	N	A	TA
1) Esperé poco tiempo para que me atendieran.	1	2	3	4	5
2) El servicio se inició de inmediato cuando llegué.	1	2	3	4	5
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
n) Estoy satisfecho con la atención que brindaron.	1	2	3	4	5

Es usted cliente de nuestra empresa: Si \_\_\_\_ No \_\_\_\_

Con qué frecuencia semanal visita nuestra empresa: \_\_\_\_\_

Observar que en el ejemplo, las variables sexo, edad se ubican como datos de identificación, la consulta de ser cliente de la empresa y la frecuencia semanal que se ubican al final, se pueden considerar como datos de clasificación pudiendo ubicar en la sección de datos de identificación.

Observar que la solicitud de cooperación y las instrucciones, se han redactado en el párrafo después de la edad y la información solicitada se asocia a las aseveraciones 1 a n.

### Me ejercito:

Elaboro un ejemplo de cuestionario conforme a la figura 10, me preparo para compartirlo con mis compañeros.

## b. Diseño del cuestionario

El diseño de un cuestionario es más un arte que una tarea científica. No existen pasos, principios o pautas que garanticen un cuestionario efectivo y eficiente. El diseño de un cuestionario es una técnica aprendida por el investigador a través de la experiencia y no por medio de la lectura de una serie de pautas. La única forma de desarrollar esta habilidad es elaborar un cuestionario, utilizarlo en una serie de entrevistas, analizar sus deficiencias y revisarlo.

Lo que sabemos acerca del diseño de un cuestionario se deriva de la experiencia de los investigadores que se han especializado en esta área. De esta experiencia acumulada han surgido una serie de reglas o pautas que pueden ser de gran utilidad para el investigador principiante enfrentado a la tarea de diseñar un cuestionario. Aunque estas reglas son útiles para evitar errores graves, el toque final del diseño del cuestionario surge

de la inspiración creativa del investigador capacitado. Por consiguiente, aunque las pautas analizadas en esta sección dan las bases para el diseño del cuestionario, en última instancia, la calidad del mismo depende de la habilidad y criterio del investigador, de un conocimiento claro de la información requerida, de una gran sensibilidad con respecto al papel del encuestado y de una exhaustiva serie de prepruebas.

El análisis del diseño del cuestionario se organizará en una secuencia de siete pasos. En cada uno de ellos se presentarán diferentes pautas para la elaboración de cuestionarios.

Aunque estas reglas se ofrecen como parte de un enfoque paso por paso para el desarrollo de un cuestionario, en la práctica los pasos están muy interrelacionados. Las decisiones que se toman con anterioridad a la secuencia, frecuentemente influyen en las alternativas posteriores de la misma y viceversa. Los siete pasos tal como se presentan en la figura 11 son:

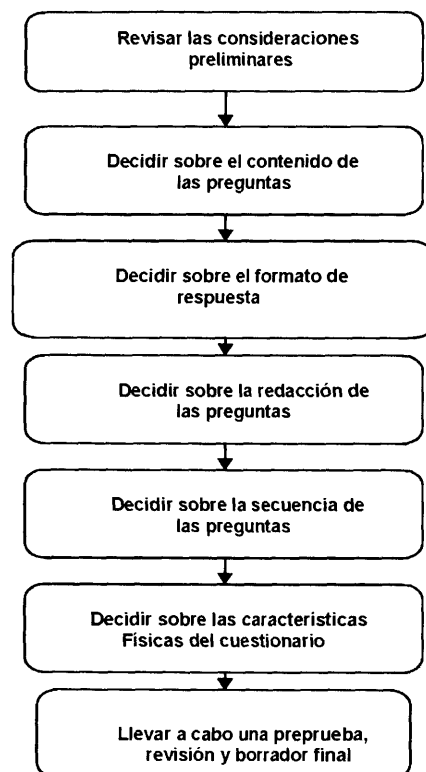


Figura 11

Fuente: Thoma C. Kinnear y James R. Taylor.  
Investigación de Mercados. Un enfoque aplicado. 1994.  
Página 313.

### **c. Revisión de las consideraciones preliminares**

Las consideraciones preliminares para realizar una investigación se asocian al establecimiento de un vínculo efectivo entre el proceso de toma de decisiones y el proceso de investigación. Un aspecto central de este proceso es el desarrollo de los objetivos de investigación y la enumeración de las necesidades de información. Debe formularse el diseño de investigación, visualizar y planear los pasos en el proceso de

investigación. Por consiguiente deben tomarse muchas decisiones antes del diseño del cuestionario.

La etapa del diseño del cuestionario supone que el proyecto de investigación está muy avanzado y que se han tomado muchas decisiones. Las decisiones relacionadas con el diseño del cuestionario deben basarse en las decisiones relacionadas con otros aspectos del proyecto de investigación y ser compatibles con ellas.

Las decisiones previas con respecto al tipo de diseño de investigación y fuentes de información influyen directamente sobre la característica y función del cuestionario dentro del proyecto de investigación. Es esencial tener una visión clara de la población objetivo y conocer los detalles del plan de muestreo.

El diseño del cuestionario está muy influido por las características del grupo de encuestados. Cuanto más heterogéneo sea el grupo de encuestados, más difícil será diseñar sólo un cuestionario que sea apropiado para todos. Generalmente el cuestionario debe diseñarse para que lo pueda comprender hasta el menos hábil de los encuestados.

Es necesario especificar las escalas de medición y los medios de comunicación que se utilizarán.

***Ejemplo:***

*Retomando el ejemplo del cuestionario de satisfacción de clientes para una empresa de servicios, observemos que se ha establecido una escala de medición tipo Likert, para recolectar los datos de los clientes que se les consulte, además el cuestionario como está estructurado puede ser autoadministrado por los clientes o levantada la información por medio de encuestadores, es así que desde esta perspectiva se tiene ya establecido el medio de comunicación y la escala de medición concreta que se utilizará para la recolección de los datos.*

Se deben visualizar las etapas del procesamiento y análisis de datos, así como la naturaleza de los hallazgos de la investigación. Los aspectos tácticos del diseño de cuestionarios están estrechamente relacionados con estas etapas finales del proceso de investigación.

Por otra parte, antes de diseñar el cuestionario, el investigador debe tener una lista detallada de las necesidades de información, así como una definición clara del grupo de encuestados. El cuestionario es el vínculo entre la información que se necesita y los datos que van a recolectarse.

Lógicamente las preguntas del cuestionario deben fluir de la lista de necesidades de información. Es obvio que no debe incluirse en el cuestionario ninguna pregunta, a menos que se relacione con una necesidad específica de información. Sin embargo, en la práctica existe una fuerte tendencia a incluir preguntas que parecen "interesantes" pero que no tienen un vínculo específico con las necesidades de información. Las preguntas innecesarias aumentan el costo de la encuesta e incrementan las demandas al encuestado.

#### **d. Decidir sobre el contenido de las preguntas**

El contenido de las preguntas está influido por la habilidad y/o buena voluntad del encuestado para responder con precisión.

Habilidad para responder con precisión. Suponiendo que los datos deseados sean pertinentes al problema de decisión, el investigador debe estar consciente de la capacidad del encuestado para proporcionar los datos. Muchos tipos de datos no se pueden recolectar de manera exacta de los encuestados. Los datos inexactos pueden provenir de dos fuentes: El encuestado está desinformado y el encuestado es olvidadizo.

El encuestado esta desinformado. A menudo se nos formulan preguntas sobre temas que no conocemos. Como estudiantes, con mucha frecuencia enfrentamos esta situación. En un cuestionario las preguntas pueden poner a los encuestados en la misma situación. Es posible que se les solicite que proporcionen datos sobre el ingreso bruto mensual o acerca de las compras con tarjeta de crédito de su cónyuge cuando no poseen los datos exactos. Es posible que se les formulen preguntas sobre anuncios publicitarios, productos, marcas o tiendas minoristas que desconocen por completo.

Los investigadores han descubierto que los encuestados, como los estudiantes, con frecuencia responden preguntas aunque desconozcan el tema, quizá debido a que no desean admitir esta falta de conocimiento. Tal situación representa una grave fuente de error de medición.

Algunas veces la manera como se formula la pregunta estimula al encuestado para que la responda, porque implica que debería conocer la respuesta.

Consideremos la pregunta: "¿Cuál es la tasa de interés corriente que recibe en su cuenta de ahorros?" Esto implica que el encuestado debe saber la respuesta. Una alternativa podría ser: "¿Conoce la tasa de interés corriente de su cuenta de ahorros?" Esta pregunta implica que algunas personas desconocen la tasa de interés, lo cual facilita al encuestado admitir que desconoce la respuesta. Si el encuestado contesta esta pregunta en forma afirmativa, entonces puede formularse la primera pregunta.

El encuestado es olvidadizo. Con frecuencia se formulan preguntas que hace algún tiempo podíamos contestar pero que ahora hemos olvidado. Como alumnos enfrentamos continuamente esta situación.

Los estudios de investigación han demostrado que olvidamos rápidamente la mayor parte de las situaciones después de que las conocemos. La tasa de olvido es veloz durante los primeros días. Por consiguiente, cuanto más tiempo haya transcurrido desde el momento del suceso, mayor será la posibilidad de olvido.

La probabilidad de olvido está influida por la importancia del suceso y por la repetición del mismo. Es más fácil recordar sucesos importantes, como la primera persona que besamos, el primer hombre en la luna, el primer automóvil que compramos, etc. En contraste, ¿cuántos de nosotros recordamos la segunda persona a quien besamos, el segundo hombre que llegó a la Luna? También es más fácil recordar hechos que se repiten constantemente, por ejemplo, productos que compramos con frecuencia o avisos publicitarios que vemos una y otra vez.

Cuando las necesidades de información de un estudio requieren preguntas que solicitan al encuestado recordar sucesos sin importancia o que ocurren con poca frecuencia, el investigador se enfrenta a un problema potencialmente grave en el diseño del cuestionario.

El investigador no debe exagerar la capacidad del encuestado para recordar con precisión el suceso y las circunstancias que lo rodearon. Este es un error que puede cometerse con facilidad cuando el tema del cuestionario es particularmente importante para el investigador o para quien toma las decisiones.

Cuando nos enfrentamos con la recolección de datos acerca de sucesos sin importancia o poco frecuentes, el investigador tiene varias opciones disponibles. En primer lugar, el investigador puede tratar de entrevistar a aquellos encuestados que tengan mayores posibilidades de recordar, por ejemplo, compradores recientes. En segundo lugar, el cuestionario puede incluir técnicas que estimulen la recordación del suceso por parte del encuestado.

Muchos estudios sugieren que las preguntas que se basan en la recordación sin ayuda, preguntas que no ofrecen clave en cuanto al suceso, pueden subestimar la ocurrencia real del suceso. El enfoque basado en una recordación con ayuda es un intento para superar el problema de la memoria; este enfoque ofrece al encuestado claves con relación al suceso de interés. Los estudiantes debemos familiarizarnos con la diferencia entre recordación con y sin ayuda. Una pregunta que solicita escribir un corto ensayo sobre un tema es un ejemplo de recordación sin ayuda, mientras que una pregunta de selección múltiple es un ejemplo de recordación con ayuda.

Una pregunta sin ayuda, diseñada para medir el conocimiento de un encuestado sobre un comercial, podría ser: "¿Cuáles productos recuerda usted de los anunciados anoche en televisión?" El enfoque de recordación con ayuda enumeraría una lista de productos para el encuestado y luego preguntaría: "¿Cuáles de estos productos se anunciaron anoche en televisión?"

La ventaja de la recordación con ayuda proviene de la clave utilizada para estimular la memoria del encuestado. El grado de estimulación puede variar desde insinuaciones limitadas hasta la presentación del suceso real. Este último enfoque hacia la recordación con ayuda recibe el nombre de método de reconocimiento. Los crecientes intentos para estimular la memoria del encuestado van acompañados de la posibilidad de una sugerencia sesgada que resulta de la presentación de las claves. Cuando se emplea el enfoque de recordación con ayuda, generalmente los investigadores recomiendan que es útil tener varios niveles de estimulación sucesiva. Esto permite que el investigador analice la influencia de la secuencia de preguntas y seleccione un nivel de estimulación más apropiado para el estudio.

Deseo de responder con precisión. Suponiendo que el encuestado puede responder la pregunta con precisión, el siguiente aspecto consiste en determinar su voluntad para hacerlo.

La renuencia a responder con precisión puede reflejarse en: Un rechazo para responder una pregunta o una serie de preguntas, es decir, error de no respuesta por ítem (o pregunta) y suministrar deliberadamente una respuesta incorrecta o distorsionada a una pregunta, esto es, error de medición.



Existen tres razones para no estar dispuesto a responder con exactitud: La situación no es apropiada para revelar los datos, revelar los datos sería algo embarazoso y revelar los datos es una amenaza potencial al prestigio o a las normas de los encuestados.

Es importante recordar que el encuestado tiene una motivación limitada para responder con exactitud a las preguntas. En el caso de las entrevistas personal y telefónica, la presencia del entrevistador puede dar como resultado una mayor preocupación del encuestado hacia cómo reaccionaría el entrevistador a las respuestas y no a la exactitud de éstas. Esta situación se presenta especialmente si las preguntas son embarazosas o si representan una amenaza al prestigio y a las normas del encuestado. El resultado puede ser un error de no respuesta por ítem o, lo que es todavía peor, respuestas inexactas.

El deseo del encuestado para responder a las preguntas está condicionando por el contexto de la entrevista. Una pregunta relacionada con los hábitos de higiene personal puede ser apropiada si la formula una enfermera o un médico como parte de un examen físico, pero puede ser inapropiada si la formula un entrevistador que está realizando un estudio para un fabricante farmacéutico.

La buena voluntad del encuestado para responder a una pregunta es también una función de su comprensión sobre el propósito legítimo para el cual se necesitan los datos. La recolección de datos de clasificación puede ser un serio problema en este caso. El encuestado puede dudar sobre si debe proporcionar datos exactos cuando se enfrenta repentinamente a preguntas personales relacionadas con la edad, la ocupación y los ingresos.

Las preguntas que ponen en aprietos al encuestado o que contienen un elemento de prestigio o de adhesión a normas sociales pueden dar como resultado una respuesta sesgada. Esta situación se presenta especialmente cuando se emplea una entrevista personal o telefónica.

Los investigadores han encontrado que preguntas sobre temas como comportamiento y actitudes sexuales, número de accidentes automovilísticos o compra de productos de higiene personal y de bebidas alcohólicas pueden poner en aprietos al encuestado, y conducen a un rechazo a responder o a la distorsión de su respuesta.

Entre las preguntas que tienen un elemento de prestigio o de adhesión a normas sociales se incluyen el nivel de educación, los ingresos percibidos y la cantidad de tiempo que se dedica a leer o a ver televisión educativa. Generalmente las respuestas a este tipo de preguntas dan como resultado un sesgo ascendente en las mismas.

### **e. Decidir sobre el formato de respuestas**

Una vez que se han analizado los problemas relacionados con el contenido de las preguntas, el siguiente tema se refiere al tipo de preguntas que debe emplearse. En este caso la preocupación está asociada con el grado de estructura impuesto a las respuestas del entrevistado. Los tres tipos de preguntas varían entre formatos de respuesta no estructurados y estructurados: Preguntas de respuesta abierta, preguntas de selección múltiple y preguntas dicotómicas.

## **1) Preguntas de respuesta abierta.**

Una pregunta de respuesta abierta requiere que los encuestados proporcionen su respuesta a la pregunta. Generalmente nos referimos a esta pregunta como pregunta de libre respuesta. Por ejemplo, puede preguntarse al encuestado: "¿Cuál es su reacción a esta nueva presentación del producto X?" En la entrevista por correo debe dejarse un espacio en blanco para que el encuestado escriba la respuesta. En las entrevistas personal y telefónica, el encuestado informa verbalmente la respuesta al entrevistador, quien a su vez escribe la respuesta en el cuestionario.

### **a) Ventajas de las preguntas de respuesta abierta**

Las preguntas de respuesta abierta sirven como una excelente primera pregunta de un tema. Permiten que se expresen las actitudes generales, lo que puede ser de gran ayuda para interpretar las preguntas más estructuradas.

Además, establecen una armonía y logran obtener la cooperación del encuestado para que responda preguntas más específicas y estructuradas. Las preguntas introductorias de respuesta abierta son especialmente importantes en las encuestas por correo.

Las preguntas de respuesta abierta influyen sobre las respuestas en menor grado que las preguntas de selección múltiple o dicotómicas. Los encuestados no están influenciados por un conjunto predeterminado de alternativas de respuesta y pueden expresar libremente puntos de vista divergentes a las expectativas del investigador.

Por otro lado, las preguntas de respuesta abierta pueden proporcionar al investigador ideas, comentarios al margen y explicaciones útiles para desarrollar un "tacto" para los hallazgos de la investigación. El informe final puede incluir citas de las preguntas de respuesta abierta para dar realismo y vida a los hallazgos más estructurados de la investigación.

### **b) Desventajas de las preguntas de respuesta abierta.**

Una de las principales desventajas de las preguntas de respuesta abierta es el alto potencial de sesgo por parte del entrevistador. Los entrevistadores rara vez registran al pie de la letra las respuestas de los encuestados. Esto da como resultado que el entrevistador resuma la respuesta del encuestado o suprima aquellos aspectos de la respuesta que él considere poco importantes. Además, aquellos entrevistadores que escriben despacio o que no emplean la taquigrafía, usualmente no registran algunas partes de la respuesta debido a las limitaciones de tiempo. Cuanto más resuma y edite el entrevistador las respuestas del encuestado, mayor será la diferencia entre las respuestas registradas y las reales. Debe utilizarse una grabadora si se desea que las respuestas se registren palabra por palabra.

Una segunda desventaja de las preguntas de respuesta abierta radica en el tiempo y costo asociados con la codificación de las respuestas. En el caso de una encuesta grande, se requieren procedimientos exhaustivos de codificación para resumir las respuestas divergentes en un formato útil para el análisis y la presentación de los datos.

El tiempo y costo de este proceso de codificación puede representar una parte significativa del costo total del proyecto de investigación.

Algunas veces se utilizan preguntas precodificadas con el propósito de aprovechar las ventajas de las preguntas de respuesta abierta y disminuir algo del tiempo y costo asociados al proceso de edición y codificación. Una pregunta precodificada es una pregunta de selección múltiple que se presenta al encuestado como una pregunta de respuesta abierta. Las alternativas de respuesta no se leen al encuestado. Por el contrario, el entrevistador selecciona la alternativa de respuesta adecuada basándose en la respuesta del encuestado a la pregunta de respuesta abierta.

Este enfoque funciona bien cuando el encuestado capta fácilmente en su mente la respuesta y cuando las posibles respuestas son limitadas en su variedad, por ejemplo, el número de miembros de una familia, tiempo de uso de un refrigerador o gastos mensuales en comida.

Sin embargo, las preguntas que no aparecen bien formuladas en la mente del encuestado y que dan como resultado una variedad de respuestas, presentan una alta probabilidad de sesgo por parte del entrevistador.

Otras desventajas incluyen la ponderación extra implícita que se da a los encuestados que se expresan en forma precisa y que tienden a expresar más puntos en sus respuestas.

Asimismo, las preguntas de respuesta abierta son menos apropiadas para los cuestionarios auto administrados. Esto se debe a que los encuestados tienden a ser más breves cuando escriben que cuando hablan, además existe un problema cuando la escritura es poco legible.

Finalmente, en comparación con preguntas que tienen formatos de respuesta estructuradas, las preguntas de respuesta abierta son entre tres y cinco veces más costosas debido a la complejidad asociada con el procesamiento de los datos.

En general, las preguntas de respuesta abierta son más apropiadas para la investigación exploratoria y la investigación diseñada para desarrollar preguntas más estructuradas. Aunque el costo para desarrollar preguntas estructuradas efectivas puede ser alto, debe evaluarse en contraste con las desventajas de las preguntas de respuesta abierta.

## **2) Preguntas de selección múltiple**

La pregunta de selección múltiple exige que el encuestado seleccione una respuesta de una lista suministrada en la pregunta o a continuación de la misma. Al encuestado se le puede solicitar que seleccione una o más de las alternativas que se le presentan.

### **a) Ventajas de las preguntas de selección múltiple**

Las preguntas de selección múltiple superan muchas de las desventajas asociadas con las preguntas de respuesta abierta. Lo más importante es que reducen el sesgo del

entrevistador y el costo y tiempo asociados con el procesamiento de los datos. Normalmente el entrevistador encontrará que este tipo de preguntas es más fácil y rápido de administrar. Finalmente, con cuestionarios auto administrados es difícil que los encuestados mantengan la cooperación a menos que la mayor parte de las preguntas tengan un formato de respuesta estructurado.

### **b) Desventajas de las preguntas de selección múltiple**

En oposición a estas ventajas, deben evaluarse algunas desventajas. En primer lugar, el diseño de preguntas efectivas de selección múltiple requiere de una cantidad considerable de tiempo y costo. A menudo se requiere un estudio exploratorio con base en preguntas de respuesta abierta para formular las alternativas de respuesta. Si estas últimas no incluyen una o más de las respuestas predominantes, se genera un sesgo sustancial en los resultados. Aunque existe la alternativa de "otros (especificar)", el encuestado tiende a seleccionar entre las alternativas especificadas en vez de utilizar la alternativa "otros". En segundo lugar, las preguntas de selección múltiple tienden a sesgar los datos de acuerdo con el orden en el cual se presentan las alternativas de respuesta al encuestado.

Aspectos en el diseño de preguntas de selección múltiple. Existen dos aspectos en el diseño de preguntas de selección múltiple: El número de alternativas y el sesgo por posición. El número de alternativas que se incluyen en una pregunta generalmente se ve influenciado por los dos siguientes principios. En primer lugar, las alternativas de respuestas deben ser colectivamente exhaustivas, es decir, deben incluir todas las posibles alternativas de respuesta.

La inclusión de la alternativa clasificada como "otro (favor especificar)", acompañada por un espacio en blanco para escribir la respuesta, es un intento para cumplir con este principio. Se espera que se identifique en la alternativa "otro" aquellas principales alternativas de respuesta que se excluyeron. En segundo lugar, las alternativas de respuesta deben ser mutuamente excluyentes, es decir, los encuestados deben identificar una alternativa que represente claramente su respuesta. En algunas situaciones, el investigador puede desear que el encuestado seleccione dos o más alternativas, pero las respuestas múltiples crean problemas especiales de procesamiento de datos.

Si el conjunto de alternativas de respuesta es razonablemente breve, las alternativas se pueden incluir en la pregunta. En la mayor parte de los casos existen demasiadas alternativas por incluir y, por tanto, se enumeran al final de la pregunta. Si la lista de alternativas es muy larga, éstas deben enumerarse en una tarjeta que se le entrega al encuestado para su revisión.

Otro aspecto importante se relaciona con el sesgo de posición. En el caso de una lista de números, como los precios o los números de visitas a un almacén, se establece un sesgo hacia la posición central del conjunto de números. Cuando se involucran ideas, la primera alternativa en la lista tiene una mayor oportunidad de selección. Para controlar el sesgo de posición, el investigador debe alternar el orden en el cual se enumeran las alternativas. Este procedimiento promediará el sesgo en las respuestas. Infortunadamente no es fácil rotar la mayor parte de los números, puesto que lógicamente deben aparecer como una secuencia, por ejemplo, 5, 6, 7, 8 y 9. Aunque se presenten en desorden, el encuestado vuelve a colocarlos en secuencia antes de hacer una elección.

### **3) Preguntas dicotómicas.**

Una pregunta dicotómica es una forma extrema de la pregunta de selección múltiple que permite al encuestado sólo dos respuestas, como "sí-no", "hice-no hice", "estoy de acuerdo-no estoy de acuerdo", etc. Generalmente las dos alternativas de interés se combinan con una alternativa neutral, tal como "sin opinión" o "no sé".

#### **a) Ventajas de las preguntas dicotómicas**

Las ventajas de la pregunta dicotómica son esencialmente las mismas ventajas que se mencionaron para las preguntas de selección múltiple. Los entrevistadores encuentran que las preguntas pueden administrarse con rapidez y facilidad. La posibilidad de que se presente un sesgo por parte del entrevistador es menor y las respuestas son fáciles de codificar, procesar y analizar.

#### **b) Desventajas de las preguntas dicotómicas.**

Existe el riesgo de suponer que el grupo de encuestados aborda el tema de interés en términos dicotómicos, cuando en realidad pueden existir muchos grados de sentimientos presentes o predominar la indecisión. Forzar a los encuestados a que expresen sus puntos de vista en forma dicotómica cuando no están polarizados, puede generar resultados que contienen un error sustancial de medición. Las preguntas dicotómicas son especialmente susceptibles al error que resulta de las palabras con las cuales se formula la pregunta.

Aspectos de diseño de la pregunta dicotómica. El aspecto fundamental se relaciona con la decisión de incluir o no una alternativa neutral de respuesta en la pregunta. Si no se incluye, el encuestado se verá forzado a seleccionar entre las dos posiciones que se le presentan. Si hay disponibilidad de una alternativa neutral, y especialmente si se le muestra al encuestado, éste puede evitar tomar una posición sobre el tema mediante la selección de la alternativa neutral.

Cuando se incluye la alternativa neutral, debe disminuir el número de no respuestas e incrementarse el número de respuestas neutrales. Si un grupo significativo de encuestados es verdaderamente neutral, la inclusión de la alternativa neutral debe incrementar la exactitud de los resultados. Sin embargo, puede surgir una fuente de sesgo cuando los encuestados que no son neutrales seleccionan la alternativa neutral por razones de conveniencia, confusión, etc. Si la proporción de encuestados verdaderamente neutral es grande, es conveniente incluir la alternativa neutral. Si se considera que la proporción de encuestados neutrales es pequeña, es mejor forzar a los encuestados a que seleccionen entre las dos posiciones de interés.

#### **f. Decisión sobre la redacción (las palabras) de las preguntas**

La parte central del cuestionario está compuesta por las preguntas formuladas. Estas preguntas representan el vínculo entre los datos y las necesidades de información del estudio. Es muy importante que el investigador y el encuestado den el mismo significado a

las preguntas formuladas. Si no es así, un grave error de medición está presente en los resultados de la investigación.

En esto no podemos llegar a creer que existe una manera "correcta" o "incorrecta" de formular las preguntas. En realidad, los datos de la encuesta se generan, en vez de recolectarlos discretamente. La manera como se recolectan los datos determina, en gran parte, el carácter de los mismos. Por consiguiente se debe estar muy consciente del efecto de la redacción de la pregunta sobre el carácter de los resultados que deben obtenerse.

Puesto que no existe una forma única y correcta de redactar una pregunta, es importante que revise claramente qué efecto puede tener una determinada redacción sobre los resultados.

La figura 12 muestra nueve principios generales que se deben considerar al diseñar la redacción de una pregunta:

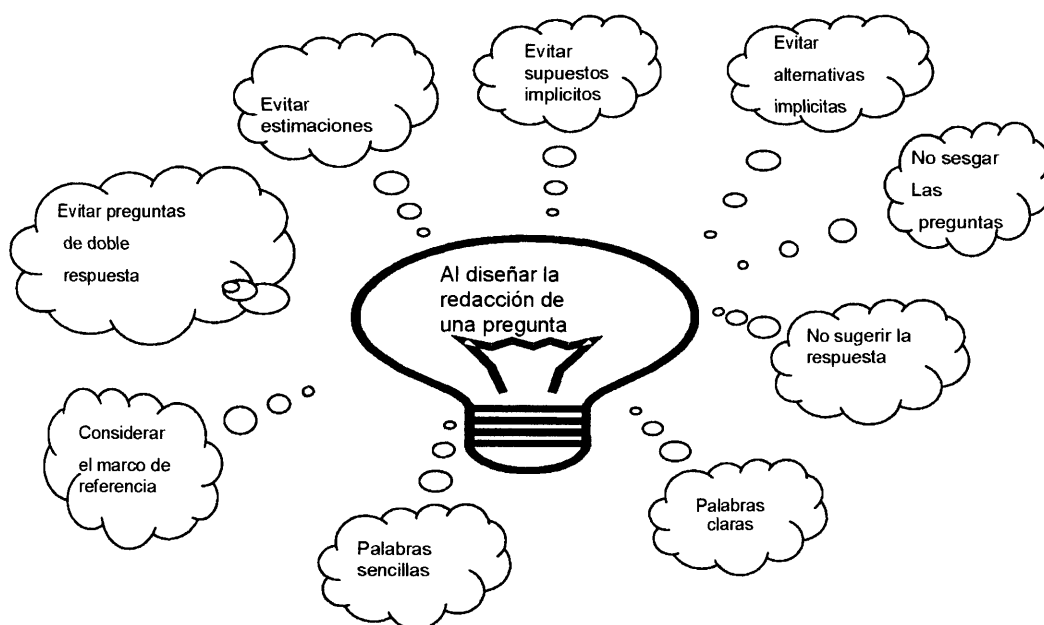


Figura 12

Las estudiaremos brevemente en el siguiente orden:

- 1) Emplear palabras sencillas,
- 2) Emplear palabras claras,
- 3) Evitar preguntas que sugieran la respuesta,
- 4) Evitar preguntas sesgadas,
- 5) Evitar alternativas implícitas,
- 6) Evitar supuestos implícitos,
- 7) Evitar estimaciones,
- 8) Evitar preguntas de doble respuesta y
- 9) Considerar el marco de referencia.

- 1) **Emplear palabras sencillas.** Las palabras que se emplean en el cuestionario deben adaptarse al nivel del vocabulario de los encuestados. Si existe alguna duda, es mejor errar hacia el lado de la sencillez.

Lógicamente, las preguntas formuladas a los niños deben enunciarse con un vocabulario más simple que las diseñadas, digamos, para los médicos. Al realizar un cuestionario para el público en general, tengamos en cuenta que muchas veces y aunque parezca sorprendente, las capacidades de vocabulario de la mayoría de los jóvenes que cursan primer año de bachillerato (con 12 años de edad) son mayores que las de muchos adultos.

- 2) **Emplee palabras claras.** Las palabras que son "claras" tienen sólo un significado conocido por todos los encuestados. Infortunadamente, identificar las palabras que sean claras o no ambiguas es más difícil de lo que uno pudiera esperar. Muchas palabras que parecen ser claras para todos pueden tener diferentes significados entre grupos de población y localizaciones geográficas.

Payne sugiere que el investigador consulte un diccionario y que formule las siguientes seis preguntas para cada una de las palabras de la pregunta:

¿Significa lo que nosotros pretendemos?

¿Tiene algún otro significado?

Si es así, ¿aclara el contexto el significado que intenta dársele?

¿Tiene la palabra más de una pronunciación?

¿Existe otra palabra con una pronunciación similar que pudiera prestarse a confusión?

¿Se sugiere una palabra o frase más sencilla?

- 3) **Evite preguntas que sugieren la respuesta.** Una pregunta que sugiere la respuesta es aquella donde se da al encuestado una clave sobre cuál debería ser la respuesta a la pregunta. Las preguntas que sugieren la respuesta, con frecuencia reflejan el punto de vista del investigador o de quien toma las decisiones con relación a la respuesta a tal pregunta. Una pregunta que sugiere la respuesta causa un error de medición constante en los hallazgos de la investigación.

Las afirmaciones diseñadas para clarificar una pregunta pueden influir en las respuestas a las preguntas que clarifican. Debemos estar conscientes de esta fuente de error de medición.

Consideremos la pregunta: "¿Posee un televisor Sony?" Esta sería una pregunta dirigida si el número informado de propietarios de televisores Sony fuera mayor que en el caso que se formulara simplemente la pregunta: "¿Qué marca de televisor tiene usted?" El uso de una marca o del nombre de una compañía en una pregunta puede hacer que el encuestado crea que esta compañía es la que patrocina la encuesta. Existe una tendencia por parte del encuestado a expresar sentimientos positivos hacia el patrocinador de la encuesta, lo cual puede concluir en un error de medición.

- 4) **Evite preguntas sesgadas.** Una pregunta sesgada incluye palabras o frases que están emocionalmente exageradas y sugieren un sentimiento de aprobación o desaprobación.

Es difícil evitar las preguntas que sugieren la respuesta, debido a que aquellas palabras o frases que sesgan a un grupo de encuestados pueden ser neutrales en otro grupo. Una preprueba del cuestionario es una forma de identificar qué grupos de encuestados consideran la pregunta sesgada.

- 5) **Evite alternativas implícitas.** Como norma es mejor expresar claramente todas las alternativas pertinentes a una pregunta, a menos que exista una razón especial para no hacerlo. Al utilizar alternativas explícitas, la investigación indica que el orden en el cual se presentan las alternativas puede afectar la respuesta. Cuando el número de alternativas es grande y complejo o muy similar en preferencia, las alternativas que aparecen al final de la lista tienen mayor posibilidad de seleccionarse.
- 6) **Evite supuestos implícitos.** Es fácil diseñar una pregunta en la cual la respuesta dependa de un número de supuestos implícitos. Consideremos la pregunta: "¿Está usted a favor de que se disminuya la cantidad de azúcar permitida en los cereales para niños?" Implícito en esta pregunta está el hecho de que esta acción dará como resultado algo favorable, por ejemplo, un índice menor de caries dental. Una mejor redacción sería: "¿Está usted a favor de que se reduzca la cantidad de azúcar permitida en el cereal para niños, si esto diera como resultado...?" El no hacer explícito el supuesto en una pregunta, con frecuencia da como resultado una sobreestimación del apoyo del encuestado al tema en cuestión.
- 7) **Evite estimaciones.** Las preguntas deben diseñarse de tal manera que el encuestado no tenga que contestarlas dando una estimación o haciendo una generalización. Consideremos la pregunta: "¿Cuántos rollos de papel higiénico compra anualmente?" Esta pregunta requiere que el encuestado determine el número de rollos de papel higiénico que compra en un mes y lo multiplique por 12. Los resultados de la encuesta serían más exactos si la pregunta fuese: "¿Cuántos rollos de papel higiénico compra mensualmente?" La cifra anual puede determinarse multiplicando por 12.
- 8) **Evite preguntas de doble respuesta.** Una pregunta de doble respuesta ocurre cuando la redacción requiere dos respuestas. Consideremos la pregunta: "¿Cuál es su evaluación sobre el manejo y la aceleración del vehículo doble tracción marca Toyota?" En este caso se han formulado dos preguntas disfrazadas en una sola. Como regla, cuando una pregunta incluye "y", revísela para comprobar si requiere dos respuestas.
- 9) **Considere el marco de referencia.** Por marco de referencia entendemos el punto de vista del encuestado al responder a la pregunta. Consideremos estas dos preguntas: "¿Los fabricantes de automóviles están logrando un progreso satisfactorio en el control de las emisiones de los automóviles?" y "¿Está usted satisfecho con el progreso que están logrando los fabricantes de automóviles en el control de las emisiones de éstos?"

El punto de vista de la primera pregunta es que una evaluación objetiva se basa en la forma como reaccionarían las personas en general a esta pregunta. La segunda pregunta está orientada al sentimiento personal del encuestado con relación al tema de las emisiones de los automóviles y es más subjetiva a este respecto. Los objetivos del estudio de investigación determinarán cuál es el marco de referencia más apropiado. El asunto es debemos estar conscientes de que la opinión del encuestado puede influir seriamente en los resultados del estudio.



### **A manera de ejemplo**

Si se tiene un grupo de preguntas a realizar y tenemos identificado que en la pregunta 5 por ejemplo que tiene dos posibilidades respuestas, si es "sí", continua en la pregunta 6, si es "no" continua en la pregunta 8, dado que las preguntas 6 y 7 no son aplicables para los entrevistados que nos responden "no".

Es así que podremos realizar un flujo de la secuencia de las preguntas, una parte de los encuestadores tendrán que contestar todas las preguntas, mientras que otra parte, en este caso los que respondan "no" en la 5ª pregunta contestaran las preguntas de 1 y 5 así como las preguntas 8 hasta concluir el cuestionario, pudiendo por lo tanto realizarse dos diagramas de flujo de la secuencia de las preguntas en correspondencia de las respuestas a ser obtenidas por los entrevistados.

### **g. Decidir sobre la secuencia de las preguntas**

Una vez que se ha determinado la redacción de las preguntas, el siguiente paso es establecer su secuencia, es decir, el orden o flujo de las preguntas en el cuestionario. La secuencia de las preguntas puede influir en la naturaleza de las respuestas de los encuestados y puede ser la causa de un grave error en los hallazgos de la encuesta. Aunque este aspecto del diseño del cuestionario se basa, en gran parte, en las capacidades de un investigador experimentado, existen varias pautas útiles para el investigador principiante.

- 1) **Utilice una pregunta introductoria sencilla e interesante.** La pregunta introductoria debe capturar inmediatamente el interés y la curiosidad del encuestado, o el encuestado dará por terminada la entrevista. Con frecuencia la pregunta introductoria no se relaciona con las necesidades de información del estudio; su único objetivo es lograr la cooperación del encuestado y establecer una relación de armonía. Al respecto, un buen comienzo puede ser una pregunta sencilla que solicite al encuestado expresar una actitud, puesto que a muchas personas les gusta expresar sus sentimientos y pueden hacerlo fácilmente. Este enfoque da a los encuestados confianza para que puedan contestar las preguntas restantes de la entrevista.
- 2) **Formule primero las preguntas generales.** Dentro de un tema, las preguntas generales deben preceder a las preguntas específicas. Analicemos las dos siguientes preguntas: "¿Qué consideraciones tiene usted en cuenta al comprar un cereal?" y "¿Al comprar un cereal, tiene en cuenta la cantidad de azúcar que contiene?" Si estas preguntas se formularan en orden inverso, el contenido de azúcar aparecería con más frecuencia en la respuesta de la primera pregunta, que si las preguntas se formularan en el orden en el cual las presentamos aquí. Por consiguiente, formular las preguntas generales primero y las preguntas específicas después reduce la posibilidad de que se presente un sesgo por secuencia.
- 3) **Coloque las preguntas poco interesantes y difíciles al final de la secuencia.** Es importante colocar las preguntas embarazosas, sensibles, complejas o monótonas al final del cuestionario. Después de establecer una armonía entre el entrevistador y el proceso de preguntas, el encuestado está menos propenso a objetar preguntas más exigentes y personales.

- 4) **Distribuya las preguntas en orden lógico.** El flujo del proceso de las preguntas debe ser lógico, desde la perspectiva del encuestado. Una secuencia de preguntas diseñada para facilitar el procesamiento de datos, o establecida desde la perspectiva del investigador, puede dar como resultado confusión, frustración e indecisión en el encuestado y puede influir seriamente en la cooperación y la armonía.

Cuando las necesidades de información de un estudio son extensivas y diferentes grupos dentro de la muestra necesitan que se les formulen diferentes preguntas, es de gran ayuda elaborar un "diagrama de flujo" de la secuencia de las preguntas. El diagrama de flujo puede ser de gran utilidad para visualizar la estructura del cuestionario y asegurar que las preguntas formuladas fluyan en forma lógica.

#### **h. Decidir sobre las características físicas del cuestionario.**

La apariencia física del cuestionario puede influir en el logro de la cooperación del encuestado. Este caso se presenta, particularmente, en las encuestas por correo. Con frecuencia la calidad del papel e impresión determinan la primera reacción del encuestado hacia el cuestionario. Es importante que el nombre de la organización patrocinadora de la encuesta (por lo general un nombre ficticio para evitar sesgo) y el nombre del proyecto aparezcan claramente en la primera página.

En el caso de las entrevistas personal y telefónica, los cuestionarios deben enumerarse en serie. Esto facilita el control del cuestionario en las operaciones de campo y durante el procesamiento de datos. Los cuestionarios por correo no tienen que identificarse numéricamente cuando es importante mantener el anonimato del encuestado.

Finalmente el formato de una pregunta puede influir en la respuesta. En el caso de los cuestionarios auto administrados y los administrados por un entrevistador, los investigadores han descubierto que cuanto más líneas o espacios en blanco se dejen para registrar la respuesta a las preguntas de respuesta abierta, más extensa será la respuesta.

#### **i. Llevar a cabo una preprueba, revisión y borrador final.**

Antes de que el cuestionario esté listo para las operaciones de campo, necesita una preprueba y revisión. La preprueba se refiere a la prueba inicial de uno o más aspectos del diseño de investigación. En este caso nos interesa la preprueba del diseño del cuestionario, porque se necesita una preprueba para buscar las áreas que deben mejorarse, a pesar de los conocimientos del investigador. La mayor parte de los cuestionarios requieren, por lo menos, una preprueba y una revisión antes de estar listos para las operaciones de campo.

Preferiblemente el cuestionario debe someterse a una preprueba de la misma forma que va a emplearse en el estudio final. Sin embargo si se tiene el borrador inicial del cuestionario, es mejor hacer una preprueba con entrevistadores personales, aunque la encuesta vaya a realizarse por correo o por teléfono. Es importante que sólo se utilicen

los mejores entrevistadores para el trabajo de preprueba. Un entrevistador capacitado puede responder a las solicitudes de una explicación, detectar las áreas de confusión y explorar la naturaleza de esta confusión. El entrevistador debe estar muy consciente de las palabras que no entiendan todos los encuestados, debe probar la secuencia de las preguntas y anotar las dificultades mecánicas y similares. Por último, será necesario someter a preprueba el cuestionario revisado, tal como va a presentarse en la encuesta final.

En la preprueba puede utilizarse el formato de respuesta abierta para determinar las categorías de respuesta apropiadas de lo que se convertirá en una pregunta de selección múltiple en el cuestionario final. Debe llevarse a cabo una nueva preprueba para descubrir cualquier problema con las categorías normalizadas de respuesta creada.

El número de personas entrevistadas en una preprueba puede oscilar entre 15 y 30. La muestra debe ser similar a la de las personas a ser entrevistadas en el estudio principal. Siempre que se hagan cambios significativos en el cuestionario, deberá realizarse otra preprueba. Si los resultados de la preprueba sugieren cambios menores, el cuestionario está listo para el borrador final y para su distribución a los operadores de campo.

### **Actividad de Autoaprendizaje No. 4**

1. Explico cual es la función de los objetivos en una investigación.
2. Valoro la importancia de la encuesta por muestreo.
3. Escribo un ejemplo en el que se identifique los componentes del cuestionario.
4. Justifico cuales son las ventajas y desventajas de las preguntas abiertas
5. Explico a qué se le denomina pregunta sesgada y cuales son sus resultados

Luego busco las respuestas a esta actividad en las páginas No. 117 y 118, al final de la unidad autoformativa I y me retroalimento.

## **E. ORIENTACIONES PARA LA ELABORACIÓN DEL TRABAJO FINAL**

---

En esta parte revisaremos las orientaciones que le servirán para la elaboración del trabajo final, recomendamos seguir paso a paso cada una de las etapas que se describen y cumpliendo con los tiempos que se establecen para cada etapa.

### **1. Generalidades**

El trabajo final es un estudio por medio de una encuesta por muestreo que se realizará sobre una determinada población. El trabajo final es un proceso de aplicación integral del aprendizaje obtenido en el módulo hacia una realidad concreta de interés del estudiante. En él intervienen tanto los estudiantes, el tutor, el material autoformativo y el contexto en el cuál vive el estudiante. Para el desarrollo del trabajo final diferenciaremos tres momentos o etapas que brevemente detallamos a continuación:

#### **a. Primera etapa: Elaboración de la propuesta del Plan del Trabajo Final.**

En esta primera etapa se deberá haber aprendido completamente la primera unidad autoformativa y se elaborará conforme los pasos que se estudian en el tema D de la presente unidad autoformativa: "*Planeación de una encuesta por muestreo*". Nos podremos auxiliar para su elaboración en las sugerencias que en esta parte se orientan como complemento a lo estudiado en el cuarto tema de la presente unidad autoformativa.

La propuesta del plan deberá elaborarse entre la primera y segunda sesión tutorial.

#### **b. Segunda etapa: Entrega de la propuesta y revisión por el tutor.**

En esta segunda etapa, el estudiante deberá entregar su propuesta de plan en la segunda sesión tutorial.

El tutor revisará las propuestas y realizará las observaciones y sugerencias si el caso lo amerita, para que sean incluidas en el plan que servirá de guía para ejecutar el trabajo final.

El tutor regresará a los estudiantes la propuesta con sus observaciones y sugerencias en la tercera sesión tutorial.

#### **c. Tercera etapa: Ajuste y aplicación del plan, entrega y breve presentación del trabajo final.**

En esta etapa se deberán integrar las observaciones y sugerencias realizadas por el tutor al plan de trabajo, pasando inmediatamente a la aplicación – ejecución del plan,

obteniéndose una vez concluido el estudio de la población por medio de la encuesta por muestreo diseñada en la primera etapa.

En la última sesión tutorial (quinta), los trabajos serán entregados al tutor para su evaluación, debiendo estar preparados para realizar una breve exposición de los aspectos más relevantes del estudio realizado.

La entrega al tutor del trabajo final se realizará en una copia impresa, respaldada con toda la información y el trabajo final en un disquete.

## 2. Sugerencias para la preparación e implementación del plan:

Otra forma de ver la preparación e implementación del plan y que retoma lo estudiado en el tema D, de la unidad autoformativa I, es representado esquemáticamente en la figura 13.

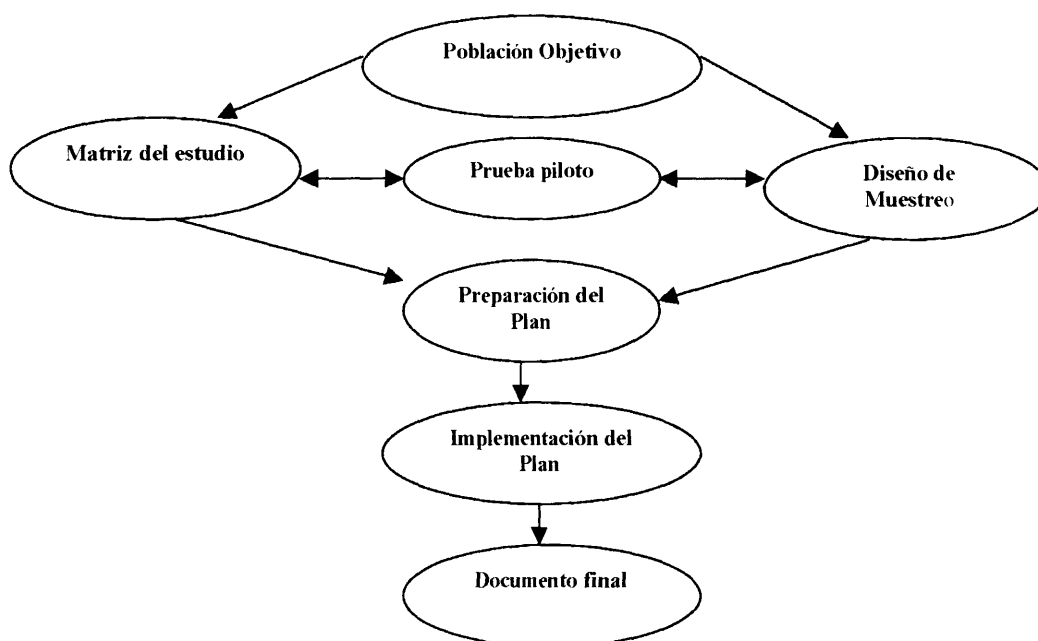


Figura 13

Observar en la figura 13 que para preparar el plan se deberá desarrollar lo concerniente a la población objetivo primeramente, posteriormente y sin importar el orden, la Matriz del Estudio o el Diseño de Muestreo y una vez realizados ambos, se procederá a la Prueba Piloto que permitirá realimentar tanto a la Matriz del Estudio como al Diseño de muestreo. Una vez realizados estos tres elementos, en conjunto serán los que nos permitirán preparar el plan. Una vez preparado el plan se implementará sirviendo de guía, para por último obtener el documento final del estudio.

## a. Población Objetivo

Ya hemos definido la población objetivo como el conjunto de elementos que posee una o varias características de interés. Observemos esa población objetivo para que podamos de ella extraer lo que nos propondremos estudiar.

### **Por ejemplo:**

*Hemos identificado como población objetivo "los estudiantes de primer año de la carrera de Administración de Empresas de la UCA" podemos:*

*Hacernos preguntas sobre la población, como por ejemplo: ¿Qué edad tienen los estudiantes? ¿Qué rendimiento Académico tienen? ¿Cuál es su procedencia respecto a Departamento geográfico? ¿Qué proporción son damas y que son caballeros?, etc.*

*Identificar algo que nos inquieta en la población: Este aspecto se relaciona con el hacernos preguntas, es así que podemos identificar estudiantes jóvenes, bajo rendimiento académico, muchos estudiantes de los departamentos, muchas damas o caballeros, etc.*

*De todas las inquietudes y preguntas que nos hagamos, seleccionaremos aquellas o aquella que más nos llama la atención. Una vez identificado el aspecto que más nos llama la atención, por ejemplo el bajo rendimiento académico, realicemos una breve valoración de que elementos causales hacen que se presente la situación de nuestro interés para el estudio.*

*Es así que podremos pensar por ejemplo que el bajo rendimiento académico es provocado por factores como: poco tiempo que se le dedica al estudio, formas en como se estudia, indisciplina en el estudio entre otros; todos estos factores los podemos resumir en los "hábitos de estudio".*

*Asimismo pueden estar involucrados muchos más factores tales como: inadecuado material del estudio, metodología del docente, recargo de asignaturas, condiciones del estudio, material de apoyo para el estudio, etc.*

De todos los factores que identifiquemos en la Breve valoración de los elementos causales que conducen a que se presente la situación, podemos elegir o relacionar un factor o elemento que influye de mayor manera en la situación que estamos interesados en estudiar en la población, esto es lo que en el módulo de técnicas de investigación llamaremos Hipótesis.

### **En nuestro ejemplo:**

*El factor que podríamos relacionar con el bajo rendimiento académico son "los hábitos del estudio" a manera de ejemplo.*

Observemos que hasta aquí lo que hemos realizado es la elección de la población objetivo, así como un tema tentativo de estudio a partir de identificar una situación de interés en la población objetivo y de valorar que elementos o factores causales me provocan la situación de interés para el estudio. Esto perfectamente podrá ser nuestro tema de estudio: "Influencia de los hábitos de estudio en el Rendimiento Académico de los estudiantes de primer año de la carrera de Administración de Empresas de la UCA".

Pero el procedimiento, no necesariamente se realiza en el orden que lo hemos señalado, puede generarse a partir de que nos preguntemos algo que vemos u oímos y que sucede a nuestro alrededor en nuestras vidas cotidianas.

***Por ejemplo nos podemos preguntar:***

*¿Por qué dan mal servicio muchas unidades de transporte colectivo de Managua? Y de esta pregunta que trata de establecer una característica en una población concreta (en este caso las unidades de transporte colectivo de la ciudad de Managua: buses, taxis, etc). Podremos entonces relacionar la situación, la población objetivo, valorar las causas de que se presente esa situación, relacionar la causa con la situación y así hasta llegar a obtener un tema de estudio concreto.*

Para los efectos de no complicar el estudio y los costos en que se pueda incurrir, se sugiere que la población objetivo que se seleccione, este en el rango de mayor o igual que 100 y menor o igual a 250 elementos.

Para complementar lo relacionado con la población, favor revisar el tema B de la presente unidad autoformativa.

Hasta aquí tenemos cubierto el paso 2 y parcialmente el 1 de la planeación de una encuesta por muestreo (tema D de la presente unidad autoformativa).

### b. Matriz del Estudio

La matriz del estudio pretende simplificar y ordenar en un cuadro todo coherente los principales elementos que nos permitirán el diseño del instrumento de medición a partir del objetivo general del estudio, está constituida por ocho columnas como se muestra en la figura 14

[illegible]

Figura 14

Las columnas que contiene la matriz del estudio son:

- 1) Objetivo General
- 2) Objetivos Específicos
- 3) Variable
- 4) Tipo de Variable
- 5) Tipo de Escala de Medición
- 6) Estadígrafo a aplicar.
- 7) Pregunta
- 8) Formato de Respuesta

## 1) Objetivo General

El objetivo general es generado a partir del tema establecido en el aspecto anteriormente abordado, es así que por ejemplo del tema que habíamos elegido: "Influencia de los hábitos de estudio en el rendimiento académico de los estudiantes de primer año de la carrera de Administración de Empresas de la UCA", podremos generar el objetivo general: "Conocer el nivel de incidencia de los hábitos de estudio en el rendimiento académico de los estudiantes de primer año de administración de empresas de la UCA en el año xxxx"

Recordar que los objetivos deben redactarse con un verbo en infinitivo, deben relacionar el factor o factores que influyen sobre la situación, ligándolos. El objetivo general nos debe mostrar lo que perseguimos con el estudio. El objetivo general no debe contener lo que perseguimos con el aprendizaje, pues este ya está incluido en los objetivos del módulo autoformativo.

## 2) Objetivos Específicos

Los objetivos específicos deben desprenderse del objetivo general, estos deben redactarse con un verbo en infinitivo, deben considerarse peldaños que juntos nos permiten alcanzar el objetivo general.

Podremos asociar cada objetivo específico a cada factor o subfactor que hemos identificado en la valoración de las causas que están presentes en la situación de interés del estudio.

La redacción debe ser muy precisa, tratando de involucrar cada objetivo específico con una sola variable causal.

### ***Algunos ejemplos de objetivos específicos son:***

*Identificar las formas en que estudian los estudiantes de primer año de Administración de empresas.*

*Conocer el número de horas que le dedican al estudio los estudiantes de primer año de Administración de empresas.*

*Diferenciar en función del sexo las formas de estudio de los estudiantes de primer año de Administración de empresas.*

## 3) Variable

En esta columna y por cada objetivo específico se deberá establecer la variable que entra en juego para identificar claramente la variable que está presente en el objetivo específico.

### ***Recordemos del módulo autoformativo el concepto de variable:***

*Variable es una característica de interés acerca de cada elemento de una población o una muestra que se analiza en un estudio estadístico y puede tomar muchos valores o clasificaciones diferentes.*



**Siguiendo con nuestro ejemplo:**

De los objetivos específicos podemos precisar las variables (o características) de nuestro interés, es así como:

Objetivo específico	Variable
Identificar las formas en que estudian los estudiantes de primer año de administración de empresas	Formas de estudio
Conocer el número de horas que le dedican al estudio los estudiantes de primer año de administración de empresas	Horas de dedicación al estudio
Diferenciar en función del sexo las formas de estudio de los estudiantes de primer año en administración de empresa	Sexo

**4) Tipo de Variable**

Aquí se especificará el tipo de variable, recordar en estadística que se han establecido variables cualitativas y cuantitativas y en las cuantitativas se han establecido en discretas y continuas.

**Continuando con nuestro ejemplo:**

Valorando los posibles valores que pueden tomar las variables (esto es el dominio de definición) podremos especificar el tipo de variable de acuerdo a lo estudiado en el módulo de estadística, por lo tanto, tenemos:

Variable	Posibles valores de la variable	Tipos de variable
Formas de estudio	Individual, en grupo	Cualitativo
Horas de dedicación al estudio	0,1,2,3...horas a la semana	Cuantitativo, discreta pudiendo ser continua. ¿Por qué?
Sexo	Masculino, Femenino	Cualitativa

**5) Tipo de escala de medición**

En función de la variable y el tipo de variable se deberá establecer el tipo de escala de medición, recordar que existen cuatro tipos de escala de medición: Nominal, Ordinal, Intervalo y de Razón. Revisar el tema A de la presente unidad autoformativa.

**Continuando con nuestro ejemplo podemos tener:**

Variable	Tipos de variable	Tipos de escala de medición
Formas de estudio	Cualitativa	Nominal
Horas de dedicación al estudio	Cuantitativa discreta	Razón
Sexo	Cualitativa	Nominal

## 6) Estadígrafo a aplicar

En esta columna y en correspondencia a la variable y su tipo de variable, así como en el objetivo general, podremos de antemano establecer el estadígrafo que se le aplicará a la variable. Revisar el cuadro 1; tercera columna (Estadística permisible) del subtema 3: "El proceso de medición", del tema A de la presente unidad autoformativa.

**En nuestro ejemplo podemos tener:**

Variable	Tipos de variable	Tipos de escala de medición	Estadígrafo a aplicar
Formas de estudio	Cualitativa	Nominal	Porcentaje Moda
Horas de dedicación al estudio	Cuantitativa discreta	Razón	Rango Media Desviación estandar
Sexo	Cualitativa	Nominal	Porcentaje

## 7) Pregunta

En esta columna y en correspondencia al objetivo específico y al general, redactar la pregunta que medirá la variable correspondiente. Tener presente el subtema 2 del tema D: Diseño de formas (cuestionarios) para la recolección de los datos de los encuestados, estudiado en el cuarto tema de la presente unidad autoformativa.

**A manera de ejemplo y continuando con el completamiento de las columnas tendremos:**

Variable	Pregunta:
Formas de estudio	¿De que manera estudia usted?
Horas de dedicación al estudio	¿Cuántas horas a la semana le dedica usted al estudio?
Sexo	Sexo:

Observar que en el caso de la variable sexo, no es necesario establecerla en forma de pregunta, además se podrá ubicar en la sección correspondiente a los datos de identificación según lo estudiado en los componentes del cuestionario, subtema B del cuarto tema de la presente unidad autoformativa: Encuestas por muestreo.

## 8) Formato de respuesta

En esta columna y en correspondencia a las preguntas generadas en la columna anterior, establecer el formato de respuestas para cada pregunta que se genere en la columna anterior. Observar que a lo largo del proceso de operacionalización que hemos venido realizando, las posibles respuestas que podemos obtener ya han sido identificadas al establecer los posibles valores de la variable.

Por ejemplo:

La variable sexo tendrá dos opciones. Masculina y femenino; la variable horas de dedicación al estudio, tendrá opciones de: 0,1,2,3... horas a la semana pudiendo usted para simplificar, establecer rangos en el formato de respuestas de 0-5, 6-10, 11-15, 16-20 horas por semana.

Tener presente el subtema 3 y 4 del tema A y el subtema 2 del tema D de la presente unidad autoformativa.

La matriz del estudio se muestra en la siguiente figura que contiene todo el proceso de operacionalización o derivación que hemos venido ejemplificando.

Podemos observar de forma general como se construye la matriz del estudio, integrando los elementos que hemos venido paso a paso construyendo:

Matriz del estudio							
Objetivo general	Objetivo específico	Variable	Tipo de Variable	Tipo de escala de medición	Estadístico permisible	Pregunta	Formato de respuesta
Conocer el nivel de incidencia de los hábitos de estudio en el rendimiento académico de los estudiantes de primer año de administración de empresas de la UCA en el año xxxx.	Identificar las formas en que estudian los estudiantes de primer año de Admón. de Emp.	Formas de Estudio	Cualitativa	Nominal	Porcentaje Moda	¿De qué manera estudia?	Individual En grupo
	Conocer las horas que le dedican al estudio los estudiantes de primer año de Admón. De Empresas.	Horas de dedicación al estudio	Cuantitativa discreta	Razón	Rango Media Desviación estándar	Cuántas horas a la semana le dedica al estudio?	0 1 2 3 horas a la semana . . n
	Diferenciar en función del sexo las formas de estudio de los estudiantes de primer año de Admón. de Empresas	Sexo	Cualitativa	Nominal	Porcentaje	Sexo?	Masculino Femenino

Observemos que al completar la matriz ya hemos diseñado el instrumento de medición a partir del establecimiento de los objetivos. El instrumento de medición estará formado por las columnas: "Pregunta" y "Formato de respuesta". Recordar aplicar los componentes del cuestionario referido en el tema cuarto, subtema B de la presente unidad autoformativa.

Hasta aquí tenemos los pasos 1, 2 y 6 de la planeación de una encuesta por muestreo.

## **c. Diseño de muestreo**

En este aspecto se estarán atendiendo los pasos 3, 4 y 5 de la planeación de una encuesta por muestreo, relacionado al marco muestral, diseño de muestreo y método de medición.

### **1) Marco Muestral**

El marco muestral, se debe construir a partir de la población objetivo por medio de establecer una lista de las unidades muestrales. Como definiremos, las unidades muestrales son un conjunto disjuntos de elementos de la población pudiendo coincidir cada elemento con la unidad muestral y que sirve para elegir en su oportunidad la muestra (ver tema A de la segunda unidad autoformativa).

En el ejemplo que hemos seguido en estas orientaciones, el marco muestral lo constituye el conjunto de elementos (estudiantes) de la población que identifiquemos y enumeremos para realizar la elección de los miembros de nuestra muestra.

### **2) Diseño de muestreo**

Como lo definiremos en la segunda unidad autoformativa, el diseño de muestreo, se asocia al procedimiento de muestreo que aplicaremos para elegir los miembros del marco muestral y que serán parte de la muestra en correspondencia al tipo de muestreo probabilístico (aleatorio simple, sistemático, estratificado, por conglomerado) que seleccionemos. Ver tema B de la presente unidad autoformativa, subtema 2, inciso b.

En correspondencia al tipo de muestreo que se seleccione, se deberá calcular el tamaño de la muestra, existiendo para cada procedimiento una fórmula para encontrar el tamaño de la muestra ( $n$ ) y que depende de los tres factores que mencionamos a continuación:

El grado de confianza

El máximo error permisible

La variabilidad de la población

Los dos primeros factores deberán ser elegidos por cada quien para su estudio particular.

El tercer factor deberá medirse por medio de la prueba piloto que se aplique.

Una vez establecido el tamaño de la muestra, se deberá proceder a la elección de los miembros de la muestra conforme al tipo de muestreo que hemos seleccionado. Este aspecto es detallado en los dos últimos temas de la segunda unidad autoformativa y en los dos temas de la tercera unidad autoformativa.

Para reforzar lo relacionado al diseño de muestreo, ver el primer tema de la segunda unidad autoformativa.

### 3) Método de medición

El método de medición se asocia a lo estudiado en el tema C de la presente unidad autoformativa, debiéndose establecer el método de recopilación en nuestro estudio, así como el medio para recolectar datos por el método de comunicación abordado en el subtema 3 de dicho tercer tema.

Es así que podemos ubicar básicamente que el método de medición que aplicaremos es el de comunicación y la técnica que aplicaremos para recolectar nuestros datos es estructurado – directo, debiendo establecer el medio de comunicación de entre los tres estudiados: Entrevista Personal, Telefónica y por correo.

#### d. Prueba Piloto

La prueba piloto, como la hemos referido en el material autoformativo, no es más que una pequeña muestra (mayor o igual a 30 elementos) que debe elegirse del marco muestral previo a la elección del tamaño de la muestra en el diseño de Muestreo y una vez que se tenga el diseño del instrumento, para poder realizar con esa pequeña muestra dos objetivos importantes en la preparación de nuestro plan:

Objetivo 1: Validar el instrumento de medición que se ha diseñado, permitiéndonos retroalimentar para corregir y ajustar el diseño del instrumento. Ver tema D, subtema 2 de la presente unidad autoformativa.

Objetivo 2: Calcular la variabilidad de la muestra piloto, permitiéndonos encontrar la varianza de la muestra piloto y que con ella sustituiremos la varianza de la población para poder determinar el tamaño de la muestra (n) que necesitamos en el paso relacionado al diseño de muestreo. Ver tema A de la segunda unidad autoformativa.

Para nuestro caso elegiremos una pequeña muestra de 15 elementos elegidos al azar de la población, para que cumplan los objetivos antes señalados.

#### e. Preparación del Plan

Una vez aplicada la prueba piloto, retroalimentado y corregido el diseño del instrumento, calculado la varianza de la muestra piloto y encontrado el tamaño de la muestra, se ha completado todo lo necesario para tener por escrito el plan.

El plan se estructurará con las siguientes unidades:

Tema

Objetivo general

Objetivos específicos

Población objetivo

Marco Muestral

Diseño de Muestreo

Método de Medición

Instrumento de Medición

Estas unidades deberán redactarse de manera clara y precisa y los principales elementos que pasarán a formar parte de la redacción han sido encontrados en el proceso que se lleve a cabo de forma ordenada en los pasos antes descritos.

Como parte del plan se deberá incluir en anexo la construcción de la matriz del estudio que sirvió de apoyo para el diseño del instrumento, así como los cálculos generados en la prueba piloto y el cálculo del tamaño de la muestra.

Se podrá adicionalmente anexar un cronograma de las actividades a ser desarrolladas para la implementación del plan versus los tiempos estimados para lograrlo.

## **f. Implementación del Plan**

Este aspecto en sí, es la fase de la aplicación del plan sobre la base del cronograma de las actividades que serán desarrolladas, se debe tener en cuenta actividades como:

- 1) Selección y adiestramiento de los investigadores de campo.
- 2) La organización del trabajo de campo.
- 3) La organización del manejo de datos.
- 4) El levantamiento de la información por parte de los investigadores de campo.
- 5) El almacenamiento y procesamiento de los datos.
- 6) El análisis de los datos, etc.

## **g. Documento final**

Una vez implementado el plan, recogidos, procesados y analizados los datos; el documento final se estructurará con las siguientes unidades:

- 1) Tema
- 2) Objetivo general
- 3) Objetivos específicos
- 4) Población objetivo
- 5) Marco Muestral
- 6) Diseño de Muestreo
- 7) Método de Medición
- 8) Instrumento de Medición
- 9) Análisis de los resultados obtenidos
- 10) Conclusiones

Observar que lo nuevo que se le ha adicionado al documento final corresponde al análisis de los resultados obtenidos y las conclusiones.

Los primeros ocho contenidos son los mismos que se han obtenido en la estructuración del plan.

En el documento final se deben entonces anexar:

Matriz del estudio.

Cálculo generado en la prueba piloto y en el tamaño de la muestra.

Cronograma de las actividades.

Matriz de los datos obtenidos en la recopilación de la información.

En el aspecto relacionado al análisis de los resultados obtenidos, se deberá aplicar los gráficos y medidas estadísticas que fueron estudiadas en el módulo de estadística y que se han referido en la columna relativa a estadígrafo a aplicar en la matriz del estudio.

En el aspecto de las conclusiones se deberá integrar los aspectos más relevantes encontrados en el análisis de resultados y que tengan relación con el objetivo general y los objetivos específicos de la matriz del estudio.

## **Actividad de autoaprendizaje 5**

1. Con base al contenido del segundo tema y las orientaciones relacionadas a la elaboración del trabajo final para la presente unidad autoformativa, selecciono una población objetivo en donde identifico una situación concreta que me motive a realizar un estudio por muestreo de dicha población.
2. Elaboro con base a la población seleccionada en el punto anterior, la Matriz del Estudio conforme la figura 14 de la presente unidad autoformativa. (Sugerencia: seguir los pasos de las orientaciones para la presentación del trabajo final, brindadas en la presente unidad autoformativa).

Después de realizar esta actividad encontraré la respuesta en la página 118 que me permitirá reforzar mis conocimientos.

## **Resumen final de la Unidad Autoformativa I**

En la primera unidad autoformativa hemos estudiado varios temas relacionados con aspectos generales de la recopilación de datos. Iniciamos el primer tema, estableciendo la importancia de los datos para el conocimiento científico, así como identificando la necesidad de realizar mediciones válidas y confiables de esos datos. Nos adentramos en la medición de las actitudes y en algunas escalas de medición de ellas. Todo esto considerado de vital importancia en las ciencias sociales y además, la base para el estudio de la Investigación de Mercados.

En nuestro segundo tema retomamos los conceptos de población y muestra, describiendo los tipos de muestreo a los que nos podremos enfrentar. Recalcamos la importancia del muestreo probabilístico, dado que al establecerse muestras representativas, nos permite realizar inferencias sobre la población objeto del estudio.

En el tercer tema de la unidad autoformativa, estudiamos los métodos de recopilación de los datos y establecemos claramente la existencia de dos tipos de métodos: observación y comunicación, estudiando sus ventajas y desventajas. Profundizamos en los medios que podemos utilizar para obtener los datos de los encuestados puntualizando las ventajas y desventajas de las entrevistas personal, telefónica y por correo.

En el cuarto tema, hemos centrado los esfuerzos en las encuestas por muestreo, habiendo estudiado los pasos por seguir en la planeación de una encuesta por muestreo y en el diseño de los cuestionarios, temas estos que revisten gran importancia en el estudio de las técnicas de recopilación de información y que nos preparan para la elaboración del plan de Trabajo Final.

Concluimos la primera unidad autoformativa con una guía práctica que le ayudará en la elaboración de la propuesta del plan del Trabajo final de la asignatura y que trata de hacer posible la integración práctica de todo el material estudiado.



Evaluación Final de la Unidad Autoformativa I

El trabajo que se anexa, ha sido elaborado por un grupo de estudiantes de Técnicas de Recopilación de la Información que estudian Licenciatura en Administración de Empresas modalidad regular de la UCA.

Revisar dicho trabajo y con base a lo estudiado por usted en la primera unidad autoformativa sugerir las mejoras que le recomienda al grupo, teniendo presente la planeación de una encuesta por muestreo que se muestra en la figura 9 de la presente unidad autoformativa y la matriz del estudio que hemos estudiado en las orientaciones para la elaboración del trabajo final.

**Población Objetivo:** Estudiantes de la carrera: Licenciatura en Administración de Empresas que estudian la asignatura de Organización II en el periodo: III cuatrimestre del 2002.

Matriz del estudio:					
Objetivo General	Objetivo Especifico	Variable	Tipo de Variable	Estadigrafo permisible	Pregunta
Analizar el nivel de interes que los jovenes tienen por informarse sobre lo que acontese en el pais	1. Identificar a quienes les interesa mas estar informados sobre las diferentes situaciones que atrviesa el pais, tomando en cuenta la edad y el sexo de los jovenes.	1.Sexo	1.Cualitativa nominal	1.Porcentaje y moda.	1. Sexo
		2. Edad	2. Cuantitativa ordinal.	2.Desv Esta. correlacion coeficiente.	2. Edad
	3. Establecer el grado de informacion que poseen los jovenes provenientes de los diferentes departamentos.	3.Lugar de origen	3.Cuantitativa nominal	3.Porcentaje moda	3. Lugar de origen.
	4. Conocer como influye el entorno familiar de los jovenes para que estos tengan interes en mantenerse informados.	4.Influencia del entorno familiar para estar inform.	4.Cualitativa nominal	4.Porcentaje moda	4. En una escala categori- ca del 1 a 5 con que frecuencia en su hogar escuchan o ven noticias.
	5. Identificar el medio de informacion que los jovenes prefieren utilizar con mayor frecuencia.	Medio de co municacion que utilizan con mayor frecuencia.	Cualitativa nominal	Porcentaje moda	Cual es el medio de co- municacion que utiliza con mayor frecuencia.
	6. Identifiar y valorar los problemas que los jovenes consideran como los mas graves del pais.	Problemas que los jove nes conside- ran como los mas graves del pais.	Cualitativa nominal	Porcentaje moda	Que tan importante es para usted informarse sobre los problemas que enfrenta el país?. Cual de los siguientes problemas considera usted que afecta mas al desarro- llo de nuestro pais?.
	7. Conocer el grado de in- formacion que poseen los estudiantes del 1° año de Lic. en Admon. de Empre. de la UCA sobre los pro- blemas de corrupcion que actualmente afronta el pais.	Grado de in- formacion que poseen los alumnos.	Cualitativa nominal.	Porcentaje moda	Como cree uste que surge la lucha contra la corrup- ción ? Esta de acuerdo con las acciones que ha tomado el gob. actual en contra de la corrupcion.
	8. Conocer las opiniones de los jovenes sobre la privatizacion de las empresas publicas como solución a los problema que enfrenta el pais.	Opinion de los jovenes.	Cualitativa nominal	Porcentil y moda	Cree usted que la priva- tización de las empresas publicas es una salida viable a la situacion del pais.

**UNIVERSIDAD CENTROAMERICANA**  
**Facultad de Ciencias Económicas y Empresariales**

*Estimado Estudiante:*

La presente encuesta tiene como propósito recolectar información de usted, para conocer cual es su nivel de interés por informarse sobre lo que acontece en nuestro país.

Favor para cada pregunta, marcar con una "X" una sola opción que vaya de acuerdo con su criterio

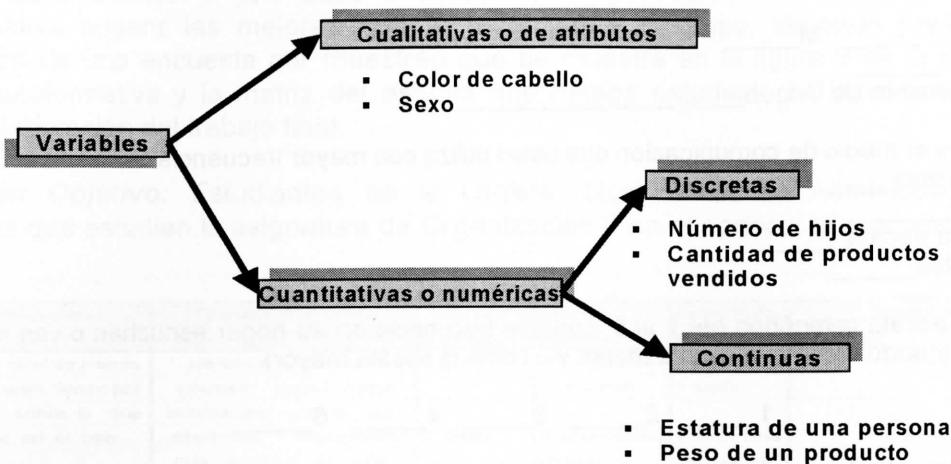
1. Sexo: F \_\_\_\_\_ M \_\_\_\_\_
2. Edad: \_\_\_\_\_
3. Departamento de Origen \_\_\_\_\_.
4. ¿Cuál es el medio de comunicación que usted utiliza con mayor frecuencia?
  - a) Televisión \_\_\_\_\_
  - b) Radio \_\_\_\_\_
  - c) Prensa escrita \_\_\_\_\_
  - d) Revistas \_\_\_\_\_
5. En una escala categórica del 1 al 5, con que frecuencia en su hogar escuchan o ven noticias (Considerando 1 como la escala menor y 5 como la escala mayor).  
  
1 \_\_\_\_\_ 2 \_\_\_\_\_ 3 \_\_\_\_\_ 4 \_\_\_\_\_ 5 \_\_\_\_\_
6. Considera usted que informarse sobre los problemas que enfrenta el país es:
  - a) Muy importante \_\_\_\_\_
  - b) Importante \_\_\_\_\_
  - c) Poco importante \_\_\_\_\_
  - d) Nada importante \_\_\_\_\_
7. ¿Cuál de los siguientes problemas considera usted que afecta mas al desarrollo de nuestro país?
  - a) Problemas culturales \_\_\_\_\_
  - b) Problemas económicos \_\_\_\_\_
  - c) Problemas políticos \_\_\_\_\_
8. Cree usted que la lucha contra la corrupción surge como:
  - a) Influencia del gobierno Norteamericano \_\_\_\_\_
  - b) Iniciativa propia del actual presidente de la república \_\_\_\_\_
  - c) Las dos anteriores \_\_\_\_\_
  - d) Presión de organismos de ayuda internacional e inversionistas \_\_\_\_\_
  - e) a, b y d \_\_\_\_\_
9. Usted esta de acuerdo con las acciones que ha tomado el actual gobierno en contra de la corrupción.  
Si \_\_\_\_\_ No \_\_\_\_\_
10. ¿Cree usted que la privatización de las empresas publicas es una salida viable a la situación el país?
  - a) Totalmente de acuerdo \_\_\_\_\_
  - b) En desacuerdo \_\_\_\_\_
  - c) Existen otras soluciones \_\_\_\_\_

***Muchas Gracias por su colaboración***

## Hoja de respuestas

### Solución a la evaluación diagnóstica

1. Realice un esquema donde se reflejen los tipos de variables estudiadas en estadística y brinde dos ejemplos de cada una de ellas.



2. Indique la tipología y el dominio de las siguientes variables: (Nota: El dominio es el conjunto de valores que puede tomar la variable, también se le conoce con el nombre de soporte).
- Número de huevos puestos por gallinas cada día en una granja. Cuantitativa discreta, Dominio: Números naturales, incluye el cero.
  - Temperatura registrada cada media hora en un observatorio. Cuantitativa continua, Dominio: Números reales.
  - Medida – grande, mediana o pequeña – de las hojas de una determinada especie de planta. Cualitativa, Dominio: {grande, mediana, pequeña}
  - Número de habitantes de los municipios de Managua. Cuantitativa discreta, Dominio: Números naturales, incluye el cero.
  - Número de hijos de una familia. Cuantitativa discreta, Dominio: Números naturales, incluye el cero.
  - Edad de los alumnos de Técnicas de Recopilación. Cuantitativa discreta, Dominio: Números naturales
  - Nivel de infección de una planta por un determinado tipo de hongo. Cualitativa, Dominio: {alta, media alta, moderada, baja, nula}
  - Número de cajillas de gaseosa producida por Kola Shaler. Cuantitativa discreta, Dominio: Números naturales, incluye el cero.
  - Percepción del servicio de transporte público en Managua. Cualitativa, Dominio: {excelente, muy bueno, bueno, regular, malo, muy malo}
  - Color de los ojos de los/las estudiantes. Cualitativa, Dominio: { azul, negro, café, verde, gris, amarillo}
3. Realice un esquema donde se reflejen los tipos de escalas o niveles de medición estudiados en estadísticas y brinde un ejemplo de cada una de ellas.
- Niveles de medición: Nominal, Ordinal, Intervalo y Razón o cociente

- Nominal: Religión que profesa: Religión Católica, Evangélica, Protestante, etc.
  - Ordinal: Rango de calificaciones de los estudiantes: Excelente, Muy bueno, Bueno, Regular y Malo.
  - Intervalo: Escala de temperatura Fahrenheit.
  - Razón: Salario devengado al año.
4. Anote el nivel de medición de las variables que se presentan en el punto número 2.
- Número de huevos puestos por gallinas cada día en una granja. Razón.
  - Temperatura registrada cada media hora en un observatorio. Intervalo.
  - Medida – grande, mediana o pequeña – de las hojas de una determinada especie de planta. Ordinal.
  - Número de habitantes de los municipios de Managua. Razón
  - Número de hijos de una familia. Razón.
  - Edad de los alumnos de Técnicas de Recopilación. Razón.
  - Nivel de infección de una planta por un determinado tipo de hongo. Ordinal.
  - Número de cajillas de gaseosa producida por Kola Shaler. Razón.
  - Percepción del servicio de transporte público en Managua. Ordinal.
  - Color de los ojos de los/las estudiantes. Nominal.
5. a) calcule la media de los siguientes valores muestrales: 5, 9, 4, 10. La Media es:  $\bar{x} = 7$ , se obtiene de  $28/4$ .
- b) Demuestre con los datos obtenidos en a), que  $\sum (x - \bar{x}) = 0$ .  $(5-7)+(9-7)+(4-7)+(10-7)=0$
6. Una empresa de servicios de energía eléctrica seleccionó 20 clientes residenciales. Los siguientes son los importes en córdobas que se cargó a los clientes por el servicio eléctrico el último mes:

54	48	58	50	25	47	75	46	60	70
67	68	39	35	56	66	33	62	65	67

- a) Calcule el importe mensual promedio. C\$ 54.55, obteniéndose de  $1,091/20$
- b) ¿Es éste un dato estadístico muestral o un parámetro poblacional? Es un estadístico muestral si la empresa de servicios eléctricos sirve a más de 20 clientes.
7. Las edades de una muestra de pasajeros que viajan en el Expreso de Matagalpa a Managua fueron: 32, 21, 60, 47, 54, 17, 72, 55, 53, y 41.
- a) Calcule la amplitud total. 55, que se obtiene de  $72 - 17$
- b) Calcule la desviación media. 13.964, que se obtiene de  $139.6/10$
- c) Calcule la desviación estándar. 17.4725, que se obtiene de:  $\sqrt{\frac{23,178 - \frac{(452)^2}{10}}{(10-1)}}$
- d) Anote el valor de la media. 45.2, que se obtiene de  $452/10$
- e) Anote el valor de la varianza. 305.288256, que se obtiene de  $(17.4725)^2$
8. Calcule la varianza y la desviación estándar de los siguientes pesos muestrales: 7, 9, 11, 9, y 4 gramos.

$$\bar{x} = \frac{7 + 9 + 11 + 9 + 4}{5} = \frac{40}{5} = 8, \quad n = 5.$$

	x	x - $\bar{x}$	(x - $\bar{x}$ ) <sup>2</sup>
	7	-1	1
	9	+1	1
	11	+3	9
	9	+1	1
	4	-4	16
$\Sigma$	40	10	28

Varianza: 
$$\frac{\sum (x - \bar{x})^2}{n - 1} = \frac{28}{5 - 1} = 7$$

Desviación estándar: 
$$\sqrt{\text{varianza}} = \sqrt{7} = 2.65$$

9. Antes de efectuar una encuesta a nivel nacional, se seleccionaron 40 personas para probar el cuestionario. Una pregunta acerca de si debe o no legalizarse el aborto, requiere una respuesta de sí o no.
- a) ¿Cuál es el experimento? Estudio de 40 personas acerca del aborto
  - b) Enuncie un posible evento. Por ejemplo 28 o más a favor
  - c) Diez de las 40 personas se declararon a favor de legalizar el aborto. Con base en estas respuestas muestrales, ¿cuál es la probabilidad de que una persona específica esté a favor de tal legalización?  $\frac{10}{40} = 0.25$
  - d) ¿Qué concepto de probabilidad ilustra esto? Frecuencia relativa
  - e) ¿Los eventos son por igual probables, mutuamente excluyentes y colectivamente exhaustivos? Los eventos no son igualmente probables, pero son mutuamente excluyentes

10. Calcule la media y la varianza de la siguiente distribución probabilística discreta:

x	P(x)
0	0.20
1	0.40
2	0.30
3	0.10

Media: 
$$\mu = \sum xp(x) = 0 \times 0.20 + 1 \times 0.40 + 2 \times 0.30 + 3 \times 0.10 = 1.30$$

Varianza: 
$$\sigma^2 = \sum (x - \mu)^2 p(x) = (0 - 1.3)^2 \times 0.20 + (1 - 1.3)^2 \times 0.40 + (2 - 1.3)^2 \times 0.30 + (3 - 1.3)^2 \times 0.10 = 0.81$$

11. Una industria produce cojinetes de bolos en forma automática en una máquina Tronar BBX. Para uno de los cojinetes, la media aritmética de los diámetros se determina como 20.00mm. La desviación estándar de la producción durante un largo período se calcula como 0.150mm.

- a) ¿Qué porcentaje de los cojinetes tendrán diámetros entre 20.00 y 20.27 mm?  
46.41%, resultado de  $z = \frac{(20.27 - 20.00)}{0.15} = 1.80$ . El área bajo la curva es 0.4641
- b) ¿Qué porcentaje de los cojinetes tendrán diámetros de 20.27mm o más? 3.59%, que se obtiene de  $0.5000 - 0.4641 = 0.0359$ .
- c) ¿Qué porcentaje de los cojinetes tendrán diámetros entre 19.85 mm y 20.30 mm? 81.85%, que se obtiene de  $0.3413 + 0.4772 = 0.8185$ .
- d) ¿Qué porcentaje de los cojinetes tendrán diámetros de 19.91mm. o menos? 27.43%. El área para  $z = 0.60$  tiene un valor de 0.2257; restando  $0.5000 - 0.2257 = 0.2743$ .

Nota: Este ejercicio se resuelve utilizando la distribución probabilística normal estándar.

12. Una muestra de 10 observaciones se selecciona a partir de una población normal para la cual se conoce que la desviación estándar es 5. La media muestral tiene un valor de 20.

- a) Determine el error estándar de la media. 1.581, que se obtiene

$$\text{de: } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{10}} = 1.581.$$

- b) Explique porqué se puede utilizar la fórmula:  $\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$ , para determinar el

intervalo. Como la muestra es elegida de una población distribuida normalmente y se conoce la varianza de la población  $\sigma^2 = 25$ , el tamaño de la muestra  $n = 10$ , nos permite realizar las estimaciones. Cuando no conocemos la distribución de la población y cuando desconocemos la varianza poblacional, es conveniente que el tamaño de la muestra que se seleccione sea mayor o igual a 30 para realizar estimaciones válidas de confianza de 95% aunque el tamaño de la muestra es menor que 30.

- c) Determine el intervalo de confianza de 95% para la media de la población. Nota: Este ejercicio corresponde a la estimación por intervalo con un nivel de confianza del 95%. 16.901 a 23.099; que se obtiene de:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 20 \pm 1.96 \times 1.581 = 20 \pm 3.099$$

## Respuestas a la Actividad de Autoaprendizaje No. 1

1. Justifico la importancia del proceso de medición

La importancia radica en la obtención de datos relevantes, confiables y válidos que nos permitan comprender la realidad. Debemos entonces desarrollar un proceso de medición de los fenómenos que nos permita mostrar la realidad, en tal sentido se justifica dicho proceso para poder establecer las relaciones que subyacen en la realidad y así poder generar conocimiento sobre esa realidad.

2. A qué se le denomina variable de actitud? ¿Cuáles son sus componentes?

Se le denomina variable de actitud a algún, algunos o todos los elementos del conjunto de procesos de percepción y de evaluación, que se basa en conocimiento y que se orienta a la acción o comportamiento de una persona con respecto a un objeto o fenómeno al que está enfrentado.

En general se puede establecer la existencia de tres componentes que se le asocian, estos son:

- Cognoscitivo, percepción de la persona acerca del objeto en cuestión.
- Afectivo, sentimientos de la persona hacia el objeto.
- Comportamiento, disposición de la persona para responder con su comportamiento al objeto.

3. ¿Qué diferencias existen entre las escalas de Likert y la Semántica Diferencial?

La diferencia fundamental se asocia en que la escala semántica diferencial presenta en sus extremos dos adjetivos opuestos y el entrevistado ubicará en que nivel de esos opuestos deberá encontrarse lo que se le está preguntando, mientras que la escala de Likert utiliza frases o enunciados que el entrevistado debe indicar su grado de aceptación o discrepancia por medio de las opciones de respuestas establecidas.

4. Explico el proceso que sigue la escala de comparación por pares

- Primeramente, se trata de ubicar al encuestado en lo que se va a comparar, para lo cual se presentan los dos objetos a comparar de un conjunto de objetos.
- Se le solicita a los encuestados que seleccionen un objeto con respecto al otro, con base a la característica que se está estudiando.
- Al final se establece la preferencia de un objeto respecto al otro del par comparado, estableciéndose para el preferido el 1 y para el no preferido el 0.

5. Anoto ejemplos de error sistemático y de error aleatorio.

Ejemplos individuales a ser compartidos en la sesión tutorial.

Respuestas a la Actividad de Autoaprendizaje No. 2

1. Elaboro un cuadro comparativo en el que se observen los distintos tipos de muestras estudiadas.

Nombre de la muestra:	Procedimiento general de selección de la muestra:	Características fundamentales:	Tipo de muestreo:
Accidental	Se obtiene sin ningún plan preconcebido, las unidades elegidas resultan producto de circunstancias fortuitas	No se tiene idea del error que se introduce al elegir la muestra. No es posible generalizar los hallazgos a la población.	No Probabilístico
Intencional	Las unidades se eligen de forma arbitraria con base al criterio del investigador		
Por cuotas	Se predetermina la cantidad de elementos de cada categoría que integrarán a muestra, seleccionándose las unidades de muestreo según criterios del investigador		
Aleatoria simple	Se selecciona la muestra a partir de un método donde cada elemento de la población tenga una oportunidad igual de ser incluida en la muestra	El grado de precisión que se desea en el cálculo, se puede establecer de antemano. Se conoce la probabilidad de selección de cada unidad. Se puede estimar el error aleatorio. Los hallazgos en la muestra, se pueden generalizar a la población por medio de las inferencias estadísticas	Probabilístico
Aleatoria sistemática	Se seleccionan los elementos de la muestra de los elementos de la población, dentro de un intervalo uniforme con respecto al tiempo, orden o espacio.		
Aleatoria estratificada	Los elementos de la muestra son proporcionales a su presencia en la población, siendo seleccionados sus elementos por el método de muestreo aleatorio simple.		
Aleatoria conglomerados por	Se divide la población en grupos o conglomerados y la selección de la muestra se realiza aleatoriamente de los conglomerados		



2. Busco un ejemplo en un diario u otro medio de información sobre alguna muestra aplicada en algún caso de interés social. Analicela, escriba sus observaciones sobre el tipo de muestra, conclusiones, validez, representatividad.

La respuesta es individual conforme a la información obtenida del diario. Será compartida en la sesión tutorial.

### **Respuestas a la Actividad de Autoaprendizaje No. 3**

1. Establezco las ventajas y desventajas que se aprecian en el método de observación

#### Ventajas:

- No se basa en la buena voluntad del encuestado para suministrar los datos
- Se reduce el sesgo potencial causado por el entrevistador y el proceso de entrevista, como consecuencia los datos son más exactos.
- Algunos tipos de datos, sólo pueden recolectarse por medio de la observación.

#### Desventajas:

- Incapacidad para observar algunos tipos de aspectos que se producen en la mente humana.
- Los patrones de comportamiento observados deben ser de corta duración, ocurrir con frecuencia o ser razonablemente predecibles para que concuerden los requisitos de costos y tiempo para la recopilación de los datos.

2. Expreso mi opinión sobre la entrevista a profundidad, caracterizándola previamente.  
¿Qué habilidades debe poseer el que entrevista?

Es una entrevista personal, no estructurada que utiliza la indagación exhaustiva para lograr que un solo encuestado exprese en forma detallada sus percepciones y sentimientos sobre una situación determinada.

En tal sentido, la entrevista a profundidad es una técnica que puede ser utilizada en conjunto con otras para complementar y profundizar sobre aspectos que otras técnicas no pueden profundizar.

#### Las habilidades del que entrevista:

- Propiciador de un ambiente donde el encuestado se sienta relajado y libre para presentar sus sentimientos y percepciones.
- Hábil en la identificación de las áreas generales de discusión para luego estimular al entrevistado a que trate libremente y en profundidad lo que se está estudiando.
- Oportuno y con inteligencia para indagar sobre las respuestas que son de interés, formulando preguntas orientadas hacia ese objetivo.
- Prudente, no debiéndose involucrar de forma activa en la discusión, sino más bien debe ser un orientador para la búsqueda de la comprensión de lo abordado.
- Responsable

3. Señalo las ventajas y desventajas que posee el cuestionario (encuesta) como parte de las técnicas estructurada- directa que hemos estudiado.

Ventajas:

- Simplicidad para administrarlo
- Facilidad con que puede procesarse, analizarse e interpretarse los datos obtenidos con esta técnica.
- Permite en alguna medida controlar sesgos en las respuestas, incrementando con ello la confiabilidad de los datos.

Desventajas:

- Los encuestados pueden no estar en capacidad de proporcionar los datos deseados o que no deseen suministrarlos.
- El proceso de interrogación puede sesgar las respuestas.
- Las preguntas estructuradas y las alternativas de respuestas fijas pueden generar una pérdida en la validez de algunos tipos de datos.

#### **Respuestas a la Actividad de Autoaprendizaje No. 4**

1. Explico cual es la función de los objetivos en una investigación.

Los objetivos en una investigación cumplen con la función de orientar el esfuerzo conjunto, son la guía que nos permite orientar todo el proceso. Sin ellos, no tendríamos la certeza de que llevemos a feliz término lo que nos hemos propuesto.

2. Valoro la importancia de la encuesta por muestreo.

La importancia fundamental se asocia a la realización de inferencias acerca de la población que se estudia y con base a los resultados obtenidos en una muestra se puede estimar los parámetros poblacionales que presentan la población objeto de nuestro estudio.

3. Escribo un ejemplo en el que se identifique los componentes del cuestionario.

- Los ejemplos deben claramente identificar los componentes siguientes:
- Datos de identificación
- Solicitud de cooperación
- Instrucciones
- Información solicitada
- Datos de clasificación
- Las respuestas variarán conforme a cada quien.

4. Justifico cuales son las ventajas y desventajas de las preguntas abiertas

Ventajas:

- Permiten que se expresen las actitudes generales del entrevistado, lo que puede ser de gran ayuda para interpretar las preguntas más estructuradas.

- Establecen una armonía y pueden lograr la cooperación del encuestado para que responda preguntas más específicas y estructuradas.
- Influyen sobre las respuestas en menor grado que las preguntas de selección múltiple o dicotómicas.
- Pueden expresar libremente puntos de vista divergentes a las expectativas del investigador.
- Pueden proporcionar ideas, comentarios al margen y explicaciones útiles al investigador

**Desventajas:**

- Alto potencial de sesgo por parte del entrevistador, al no integrar completamente lo respondido por el entrevistado.
- Tiempo y costo alto relativamente, asociado con la codificación de las respuestas.
- Son menos apropiadas para los cuestionarios auto administrados.

5. Explico a qué se le denomina pregunta sesgada y cuales son sus resultados

Se le denomina pregunta sesgada a aquellas preguntas que incluyen palabras o frases que están emocionalmente exageradas y que sugieren un sentimiento de aprobación o desaprobación, generando con ello una influencia hacia la sugerencia de la respuesta y por lo tanto un error sistemático en una dirección determinada, provocando con ello una disminución en la validez de lo que se investiga.

## Respuestas a la Actividad de autoaprendizaje No. 5

1. Con base al contenido del segundo tema y las orientaciones para la elaboración del trabajo final para la presente unidad autoformativa, selecciono una población objetivo en donde identifique una situación concreta que me motive a realizar un estudio por muestreo de dicha población.

Población seleccionada de manera individual por cada estudiante.

2. Elabore con base a la población seleccionada en el punto anterior, la Matriz del Estudio conforme la figura 14 de la presente unidad autoformativa. (Sugerencia: seguir los pasos de las orientaciones para la presentación del trabajo final brindadas en la presente unidad autoformativa).

[illegible]

## Glosario de términos

**Actitud:** Conjunto de procesos de percepción permanente y de evaluación de un individuo, basado en conocimientos y orientado a la acción con respecto a un objeto o fenómeno.

**Análisis de datos:** Conjunto de transformaciones numéricas de los datos, encaminado a lograr que éstos sean interpretables con relación a la hipótesis de investigación. Tales transformaciones suelen ser de índole estadístico.

**Ciencia:** Tipo de conocimiento humano que se caracteriza por su objetivo expresión del conocimiento en forma de reglas de ámbito general— y su método —método científico.

**Componente afectivo:** Sentimientos de un individuo que son movilizados hacia un objeto.

**Componente cognoscitivo:** Persuasión que posee un individuo con relación a un objeto en cuestión.

**Componente de comportamiento:** Disposición de un individuo para responder con su comportamiento al objeto.

**Confiabilidad:** Medida que se refiere al grado hasta el cual, el proceso de medición está libre de errores aleatorios.

**Conocimiento:** Información que, sobre la Naturaleza y sobre sí misma, la especie humana ha ido acumulando a lo largo de su historia.

**Conocimiento científico:** Conocimiento adquirido por medio del método científico.

**Conocimiento vulgar:** Forma de conocimiento práctico que se presenta en la cultura popular, presente en la vida cotidiana.

**Constructo:** Variable definida teóricamente que no es accesible a la observación directa. Utilizado frecuentemente en las ciencias sociales. Su significado es similar al de concepto. Abstracción mental formada por la percepción de un fenómeno

**Cuestionario:** Conjunto previamente determinado de preguntas que se utiliza en el proceso de la encuesta para conocer características de una población. Las respuestas son, en general, cerradas.

**Encuesta:** Investigación destinada a conocer características de una población de sujetos a través de un conjunto de preguntas.

**Entrevista:** Acto de preguntar y registrar las respuestas de los elementos de la muestra.

**Entrevista estructurada:** Conjunto de preguntas previamente determinado. Las preguntas pueden ser abiertas o cerradas. Se pueden utilizar en las encuestas.

**Entrevista no estructurada:** Conjunto de preguntas que no está previamente determinado. Las preguntas son, en general, abiertas. No se utilizan en las encuestas.

**Entrevista personal:** Acto de formular y recoger las respuestas en la que el entrevistador está físicamente presente.

**Entrevista por correo:** Acto de enviar las preguntas y recoger las respuestas a través del correo. Sólo se usan cuestionarios.

**Entrevista por teléfono:** Acto de formular y recoger las respuestas a través del teléfono. Sólo se usan cuestionarios.

**Error aleatorio:** Diferencias observadas en las puntuaciones de la variable dependiente y que no son explicadas por ninguna de las variables bajo control.

**Error de estimación:** Diferencia absoluta entre el valor del estimador muestral y el parámetro poblacional, también se le denomina error de muestreo.

**Error de medición:** Diferencia absoluta entre el valor de lo medido y lo que en realidad se debería de medir.

**Error sistemático:** Tipo de error permanente que sesga las mediciones, se reproduce constantemente y en el mismo sentido provocado por imprecisiones de los instrumentos de medición.

**Extracción de muestra:** Procedimiento que se sigue para seleccionar los elementos que conformaran una muestra.

**Fenómenos:** Son los hechos captados por el investigador/ observador y percibidos a través de los sentidos, designando entonces un hecho percibido.

**Hecho:** Es todo aquello que se sabe o se supone que pertenece a la realidad.

**Hipótesis:** Explicación tentativa a un problema de investigación. Suele expresarse en forma condicional: «Si... entonces...».

**Marco muestral:** Es la lista de las unidades poblacionales de la cual se tomará la muestra.

**Medición:** Asignación de números a las características de los objetos o sucesos de acuerdo con ciertas reglas.

**Método de Comunicación:** Método de recolección de datos basado en la interrogación de los encuestados.

**Método científico:** Conjunto de pasos reglados que utiliza la Ciencia para la ampliación de sus conocimientos.

**Método de Observación:** Método clásico de recolección de datos en la investigación científica, que nos permite obtener información sobre el mundo que nos rodea.

**Muestreo:** Acto de obtener una muestra.

**Muestreo no probabilístico:** Muestreo en el que no se puede determinar cuál es la probabilidad de que un elemento de la población forme parte de la muestra.

**Muestreo probabilístico:** Muestreo que se selecciona de modo que cada integrante de la población en estudio tenga una probabilidad conocida, no igual a cero, de ser incluido en la muestra.

**Muestra representativa:** Grupo de elementos extraídos de una población de tal modo que tenga la misma configuración que ésta en relación a las variables que la definen.

**Pertinente:** Congruencia entre las expectativas de la realidad y la investigación que se realiza.

**Población:** Conjunto de elementos que comparten una propiedad en función de la cual se define a la misma.

**Pregunta abierta:** Pregunta a la que el sujeto puede contestar lo que le parezca.

**Pregunta cerrada:** Pregunta a la que el sujeto sólo puede contestar eligiendo una de las opciones que se le presentan.

**Pretest:** Proceso en el que se pone a prueba el cuestionario antes de realizar la encuesta.

**Proceso de medición:** Procedimiento que permite la asociación del sistema de números para representar los fenómenos del sistema empírico que se investiga.

**Representatividad:** Propiedad atribuible a una muestra cuando está configurada por las mismas variables de la población a la que pertenece. Cuando una muestra es representativa, las características observadas en la muestra se pueden generalizar a la población.

**Sesgo:** Error que se produce al hacer inferencias desde una muestra que no es representativa.

**Sesgo del entrevistador:** Modificación de la información dada por el sujeto debida a los intereses y valores del entrevistador.

**Sesgo de respuesta:** Influencia, en las respuestas a un cuestionario, debida al orden en que se contesta. Se produce cuando no hay entrevistador y el sujeto, antes de contestar, lee las preguntas que están a continuación.

**Trabajo de campo:** Parte del proceso de la encuesta en la que se lleva a cabo la localización de los elementos de la muestra, se realizan las preguntas y se registran las respuestas.

**Unidades de datos:**

**Unidad muestral:** Es el elemento base de la encuesta. Puede ser un individuo, una familia o un grupo determinado.

**Validez:** Medida que se refiere al grado en el cual el proceso de medición, está libre tanto de error aleatorio como de error sistemático.

## **Bibliografía**

Hayer Bob.E. Como medir la satisfacción del cliente, Diseño de encuestas, uso y métodos de análisis estadístico. 1999, Editorial Oxford, segunda edición, México.

Quivy R. y Campenhoudt L. V. Manual de investigaciones en Ciencias Sociales. 1992, Editorial Limusa, primera edición, México.

León O. G. y Montero I. Diseño de investigaciones.1997, Editorial M<sup>c</sup> Graw Hill, segunda edición, España.

Sheaffer R., Mendenhal W., y Ott L. Elementos de muestreo. 1987, Editorial Iberoamericana, tercera edición, México.

Kinnear T. C. y Taylor J.R. Investigación de mercados, un enfoque aplicado. 1994, Editorial M<sup>c</sup> Graw Hill, cuarta edición, Colombia.

Cohran W. G. Técnicas de muestreo. 1995, Editorial CECSA, primera edición, México.

Pozo J. I. Teorías cognitivas del aprendizaje. 1996, Editorial Morata, cuarta edición, España.

Mason D. B. y Lind D. A., Estadística para administración y economía 2000, Editorial Alfaomega, octava edición, México.

***Unidad Autoformativa II***  
***“Muestreo probabilístico aleatorio  
simple y aleatorio sistemático”***



## Presentación

*En la presente unidad autoformativa estudiaremos los dos primeros métodos de muestreo probabilístico de los cuatro previstos en el presente módulo. Estos métodos son el aleatorio simple y el aleatorio sistemático.*

*Primeramente esbozaremos la importancia que tiene el diseño de muestras, retomando algunos conceptos que ya hemos estudiado en estadística como por ejemplo la distribución normal, la estimación puntual y por intervalo que son la base para el estudio de lo que se conoce en estadística con el nombre de inferencia y que nos servirá para estudiar los métodos de muestreo antes señalados.*

*En el estudio de estos dos primeros métodos realizaremos en orden la definición del método de muestreo, posteriormente estudiaremos el procedimiento de selección de la muestra para concluir con la presentación de diversas fórmulas que nos ayudarán a realizar las estimaciones de la media poblacional, total poblacional, proporción poblacional y el tamaño de la muestra.*

*Al final de esta unidad, usted podrá identificar y aplicar las fórmulas para realizar los cálculos apropiados.*

## Objetivos de la unidad autoformativa II

1. Comprendo diversos conceptos relacionados al diseño de muestras probabilísticas.
2. Aplico adecuadamente los procedimientos de selección de muestras aleatorias simples y sistemáticas.
3. Aplico las formulas adecuadas para las estimaciones de la media poblacional, total poblacional, proporción poblacional y tamaño de muestras aleatorias simples y aleatorias sistemáticas.

## Contenido de la unidad autoformativa II

### Tema A. Diseño de muestras

1. Definiciones
2. Error de estimación en el muestreo
3. Estimaciones en el muestreo
4. Estimación Puntual
5. Estimación por intervalo
6. Solución del tamaño de la muestra

### Tema B. Muestreo Aleatorio simple

1. Definición
2. Procedimiento de selección de una muestra aleatoria simple
3. Estimaciones basadas en una muestra aleatoria simple.
4. Estimador de la media poblacional, varianza estimada de la media y limite para el error de estimación de la media
5. Estimador del total poblacional, varianza estimada del total y limite para el error de estimación del total
6. Estimador de la proporción población, varianza estimada de la proporción y limite para el error de estimación de la proporción poblacional
7. Cálculo del tamaño de la muestra para las medias, total poblacional y proporción poblacional.

### Tema C. Muestreo aleatorio sistemático

1. Definición
2. Procedimiento de selección de una muestra aleatoria sistemática
3. Definición de población aleatoria, ordenada, periódica y su efecto en la varianza y las estimaciones.
4. Estimaciones basadas en una muestra aleatoria sistemática con población aleatoria.
5. Estimador de la media poblacional, varianza estimada de la media y limite para el error de estimación de la media
6. Estimador del total poblacional, varianza estimada del total y limite para el error de estimación del total
7. Estimador de la proporción población, varianza estimada de la proporción y limite para el error de estimación de la proporción poblacional
8. Cálculo del tamaño de la muestra para las medias, total poblacional y proporción poblacional.

## A. DISEÑO DE MUESTRAS

El objetivo principal de un diseño de muestras es proporcionar indicaciones para que la selección de estas sean representativas de la población bajo estudio, proporcionándonos así una cantidad especificada de información a un costo mínimo.

Si la población bajo estudio es uniforme en las características que serán medidas, es posible que cualquier muestra nos produzca resultados aceptables. Pero si la población bajo estudio no es uniforme, el diseño de muestras juega un papel fundamental.

*Una población se denomina uniforme sobre la base de una característica, cuando la medida de esa característica nos resulta en un valor igual para todas las mediciones realizadas a los miembros de la población.*

### **Como ejemplo**

Al realizar un examen de una asignatura, todos los miembros del grupo que en este caso se constituye como nuestra población obtiene la misma nota.

Al usar muestreo es necesario tomar las precauciones necesarias para asegurar que éste se lleve a cabo en forma aleatoria y que la muestra sea una muestra aleatoria.

Una muestra aleatoria se selecciona de forma que toda combinación de "n" mediciones de la población tenga la misma probabilidad de ser seleccionada.

La ventaja principal del uso de diseños de muestreo aleatorio es que cuando las muestras son aleatorias, se conocen las probabilidades de incluir ciertas observaciones en la muestra. Por lo tanto es posible hacer afirmaciones de tipo probabilístico acerca de la población. Si las muestras se seleccionan en forma determinista (no aleatoria), las probabilidades de observar distintas mediciones muestrales son desconocidas y solamente se pueden hacer afirmaciones de tipo descriptivo en relación con la muestra y no con relación a la población.

El objetivo final de una encuesta por muestreo es realizar una inferencia acerca de la población bajo estudio, se basa en estimaciones de la media poblacional, el total poblacional, o la proporción. Para cada uno de estos estimadores, se dará una cota B para el error de estimación. La cota B, se asocia entonces al máximo valor posible de error que el investigador está dispuesto a permitir al realizar la estimación.

Comprenderemos como media poblacional, el valor medio que nos dará una medición de una variable a todos los miembros de la población.

### A manera de ejemplo

Si en una población de tamaño  $N$ , la suma de las mediciones de esas variables nos resulta por ejemplo 10432 y el valor de  $N = 126$ , entonces  $\mu = 84.81$ .

El total poblacional, será  $\tau = 10432$  y si de esa población por ejemplo existen 48 varones, la proporción poblacional  $p = 0.39$ .

Existen entonces una diversidad de variables que se necesitan medir y que nosotros conocemos, tal es el caso de: PIB de Nicaragua en el año 2003, que representa el total poblacional; el PIB per cápita anual 2003, representa la media poblacional; el porcentaje de mujeres de la población Nicaragüense en el 2003 representa una proporción poblacional.

### Me ejercito:

Identifico variables que conozco de una población y determino si le es aplicable una media, una proporción o un total poblacional. Me preparo para compartirlo con mis compañeros.

**Una inferencia** es la determinación de algo (una característica) acerca de una población con base a lo obtenido en una muestra.

**La estimación** es el cálculo de determinados estadísticos en la muestra como por ejemplo la media de edades, que nos permite inferir la media de edades en la población (parámetro poblacional).

Naturalmente, las fórmulas de estimación difieren según el diseño de muestreo que se use en la encuesta.

Es conveniente aclarar aquí que en varios de los pasos de la planeación de una encuesta por muestreo que hemos estudiado en la unidad autoformativa I, pasos como:

Definición de la población

Determinación de los datos a recoger

Métodos de medición

Organización del trabajo de campo

El diseño de muestras tiene poco o nada que ver, sin embargo como lo hemos señalado en su oportunidad, estas pueden influir en el nivel de error y sesgar el estudio que realizamos, razón por la cual es conveniente tener presente su importancia.

Ahora bien, para adentrarnos más al diseño de muestra conviene tener claras algunas definiciones que pasaremos a establecer.

## 1. Definiciones

El diseño de muestreo o diseño de encuesta especifica el método de obtención de la muestra.

El diseño no especifica la forma de recolectar o medir los datos reales. Especifica únicamente el método de recolección de los objetos que contienen la información requerida. Estos objetos se llaman elementos.

Cuando hablamos de diseños de muestreo nos estaremos refiriendo al procedimiento que seguimos para elegir de los miembros de la población los que serán parte de la muestra, es así que los diseños que estudiaremos serán los que establecíamos en la primera unidad autoformativa como estrategias del muestreo probabilístico (Aleatorio simple, sistemático, estratificado y por conglomerados).

Un elemento es un objeto del cual se toma una medición.

Los elementos pueden ocurrir individualmente o en grupos en la población. Un grupo de elementos, como una familia o una caja de bombillas, se llama unidad de muestreo.

Un elemento es pues un miembro de la población que en el proceso que se siga de muestreo es un potencial miembro para que forme parte de la muestra y que luego le realicemos la medición de la característica que estamos interesados en estudiar.

Las unidades muestrales son colecciones disjuntas de elementos de la población. En algunos casos una unidad muestral está constituida por un solo elemento.

Para seleccionar una muestra aleatoria de unidades muestrales, es necesaria una lista de todas las unidades muestrales contenidas en la población. Esta lista se llama marco muestral.

### **Por ejemplo**

*Si consideramos un distrito de Managua como nuestra población, observemos que podemos considerar a cada habitante del distrito como la unidad muestral, sin embargo, si de acuerdo al interés del estudio no nos interesa cada habitante, sino la familia, ésta será la unidad muestral, observemos que en esta situación la unidad muestral está constituida por varios elementos de una población.*

Una vez que tengamos definida la unidad muestral que se asocia al estudio que realizaremos, es conveniente que elaboremos una lista exhaustiva de todas las unidades muestrales que son contenidas en la población, sin que exista traslape de miembros, esto es que un elemento de la población no podrá pertenecer a dos o más unidades muestrales, sino que solamente a una. A este listado de unidades muestrales se le denomina marco muestral.

Un marco muestral es una lista de unidades muestrales.

Al efectuar una encuesta por muestreo, la primera tarea consiste en la identificación de las unidades muestrales y en la construcción del marco que contiene la lista de dichas unidades. De acuerdo al diseño de muestreo, se selecciona en forma aleatoria un número predeterminado de unidades muestrales de la población. La muestra está compuesta de los elementos contenidos en las unidades muestrales seleccionadas. Se puede entonces usar la información obtenida de la muestra para hacer inferencias acerca de determinadas características de la población.

### **En síntesis**

*El propósito del diseño de muestras es desarrollar métodos o procedimientos de selección de muestras y de estimación que proporcionen al menor costo posible, estimaciones con la suficiente precisión para realizar los estudios adecuados de las poblaciones.*

Estaremos interesados entonces en el diseño de muestras, en especificar el método o procedimiento para obtener la muestra, teniendo presente los elementos o las unidades muestrales para seleccionar la muestra aleatoria, que la seleccionaremos del marco muestral construido.

Observemos que el marco muestral podrá coincidir completamente o ser un subconjunto de la población objeto de estudio.

### **Ejemplo:**

*Consideremos que estamos interesados en estudiar (dos meses después de iniciado) en un curso académico determinado, la edad promedio de los estudiantes de primer ingreso en la carrera de administración de empresas en la UCA. Supongamos que la población ingresante es de 400 estudiantes.*

*La población bajo estudio está determinada por los estudiantes que cumplen la condición de ser de primer ingreso en la carrera de administración de empresas en la UCA para el respectivo año académico.*

*Cada estudiante que forme parte de la población bajo estudio se constituye en un elemento del cual se podrá tomar una medición relacionada a la característica que estamos interesados en estudiar; para este caso la edad promedio. Observemos que cada elemento en este caso es una unidad muestral (recordemos que una unidad muestral son colecciones disjuntas de elementos de la población y que en algunos casos una unidad muestral está constituida por un solo elemento).*

*Ahora bien, observemos que de todos los elementos (unidades muestrales) de la población, algunos (15 por ejemplo) se han retirado, cambiado de carrera o simplemente no llegan a clase, lo que nos resulta tener en total 385 estudiantes (unidades muestrales) al momento de realizar el estudio.*

*La construcción del marco muestral es entonces indispensable con las 385 unidades muestrales antes referida y no con los 400. Esto es porque existen 15 elementos que si bien es cierto en algún momento cumplían con la condición, al momento del estudio ya dejaron de ser parte de la población ( en otros casos podrán ser parte de la población, pero por otras circunstancias no es posible incluirla en el marco muestral).*

*Es así que el detalle de los 385 elementos (unidades muestrales) es nuestro marco muestral. Si tuviésemos que elegir por ejemplo una muestra de tamaño 36 para realizar las estimaciones e inferencias de la muestra a la población bajo estudio, entonces el procedimiento o método que nos permite seleccionar los 36 elementos de la muestra de los 385 elementos del marco muestral, es lo que denominaremos con el nombre de diseño de muestreo.*

Al inicio de la presente unidad hemos planteado que el objetivo final de una encuesta por muestreo es la realización de una inferencia sobre la población bajo estudio y nos hemos referido al error de estimación, este es intrínseco al muestreo, razón por la cual nos interesa profundizar para continuar con nuestro estudio.

## 2. Error de estimación en el muestreo

El error de estimación, se define como la diferencia absoluta entre el estimador muestral y el parámetro poblacional, simbólicamente lo podemos representar como:  $|\bar{\theta} - \theta| \leq B$   
Podemos observar que:

- $\bar{\theta}$  : Es un estimador muestral y asumiremos que cuando tenga una barrita encima o un gorrito ( $\bar{\cdot}$ ), será para nosotros un estimador muestral (realizamos cálculos con los datos obtenidos de la muestra).
- $\theta$  : En cambio es nuestro parámetro poblacional (se calcula con todos los datos de la población), observar que no le hemos ubicado ni barrita, ni gorrito.
- B: como lo habíamos referido anteriormente, es nuestra cota o error de estimación que se relaciona por medio de la diferencia absoluta (valor absoluto de la resta) entre el estimador y el parámetro que se desea estimar.

Al escoger el tamaño de muestra  $n$ , el investigador debe especificar una cota  $B$  para el error de estimación que está dispuesto a tolerar. Es decir, el investigador debe especificar el valor de  $B$  y posteriormente escoger el tamaño de muestra  $n$  de forma que  $\bar{\theta}$  y  $\theta$  difieran por una cantidad mayor que  $B$  solamente una pequeña fracción del tiempo.

### Por ejemplo

*Un auditor desearía que fueran escasas las posibilidades de que el saldo promedio de las cuentas por cobrar de una muestra de los libros de contabilidad de la empresa difiera del valor real por más de una cantidad especificada.*

*Continuando con nuestro ejemplo de la edad promedio de los estudiantes de primer ingreso en la carrera de administración de empresas en la UCA, antes de elegir el tamaño de la muestra  $n$ , debemos especificar el error de estimación  $B$ , por ejemplo si especificamos que nuestro error de estimación es  $B = 1$  año, en correspondencia al procedimiento de muestreo podremos encontrar un valor de  $n = n_1$ ; sin embargo si en vez de  $B = 1$  año, estamos interesados que nuestro error de estimación sea menor que el anterior, por ejemplo  $B = 0.5$  años, el valor del tamaño muestral será  $n = n_2$ , siendo  $n_2 > n_1$ , esto es el tamaño requerido de la muestra para el error de estimación de  $B = 0.5$  años deberá ser mayor que el tamaño de muestra requerido para un error de estimación de  $B = 1$  año.*

Como es de esperarse,  $B$  y  $n$  están relacionados inversamente; si el error de medición ( $B$ ) tolerable es menor, mayor será la cantidad de información requerida para satisfacer la cota especificada, es decir mayor será  $n$  (el tamaño de la muestra).

En las encuestas por muestreo existen tres fuentes de error:

- a. La variación aleatoria
- b. La especificación deficiente de la población
- c. La no-respuesta.

Veamos cada una de ellas:

### a. La variación aleatoria

La fuente más común es la variación aleatoria.

### **Por ejemplo**

*Supóngase que una organización de ventas al menudeo está interesada en estimar el ingreso promedio por vivienda en una comunidad determinada. Al seleccionar la muestra aleatoria de viviendas, puede ocurrir que se hayan seleccionado sólo aquellas cuyos habitantes pertenecen al estrato superior de ingresos. El sentido común y la intuición sugerirían la presencia de error de estimación si se obtiene un ingreso promedio por vivienda demasiado alto para esa comunidad. Sin embargo, si el error de estimación es moderado, puede pasar desapercibido y producir inferencias erróneas y, quizás, decisiones equivocadas. Si se obtiene un ingreso inflado pero creíble, la organización puede decidir no introducir al mercado líneas de productos de precio módico, que se consideran más adecuadas para comunidades de menor nivel de ingresos. Con base en un ingreso promedio por vivienda inflado, la empresa puede suponer erróneamente que en esa comunidad hay muy pocas familias cuyo ingreso es moderado o bajo.*

### **b. La especificación deficiente de la población**

Otra fuente de errores en las encuestas por muestreo está constituida por la especificación errónea (deficiente) de la población. Estos errores son muy comunes en las encuestas de opinión pública relacionadas con elecciones políticas. La población real en este tipo de encuestas está constituida por aquellas personas que votarán el día de las elecciones. Sin embargo, las encuestas típicas determinan la opinión de los votantes registrados, muchos de los cuales no votarán el día de las elecciones.

Los errores debidos a la especificación deficiente de la población pueden ser causados por otras fuentes, como una lista incorrecta de los elementos de la población, información incorrecta registrada en un libro de inventarios, selección errónea de los elementos de la muestra (como seleccionar al vecino cuando la persona que va a entrevistarse no está en casa), sensibilidad de la pregunta, errores en la recolección de información de la muestra debido al sesgo intencional o no intencional del entrevistador, o errores en el procesamiento de la información muestral. En la mayoría de los casos, dichas causas pueden ser controlables.

En las encuestas sobre consumo, se producen errores de especificación deficiente de la población cuando la muestra consiste únicamente en amas de casa y se excluyen las mujeres que trabajan, los hombres y los estudiantes, por su inaccesibilidad relativa.

### **c. La no-respuesta.**

Una fuente adicional de error en las encuestas por muestreo está constituida por la no-respuesta de algunos elementos de la muestra. Es común que el investigador suponga que los elementos de la muestra que responden y los que no lo hacen tienen características semejantes, lo cual raras veces ocurre. En encuestas de consumo, la no-respuesta ocurre generalmente entre la gente que trabaja y las que responden son usualmente amas de casa; en las encuestas de opinión, los que no responden (aquellos que "no opinan") son generalmente los miembros de la muestra que prefieren que las cosas permanezcan como están.

Para minimizar los errores producidos por la variación aleatoria, lo único que se puede hacer es seleccionar el diseño de muestreo adecuado. Sin embargo, en lo que se refiere a los errores producidos por la no-respuesta, el investigador puede tener un efecto mucho más directo. Es posible hacer un mayor esfuerzo al contactar a los que no responden o,



en algunos casos, los que no responden pueden ser reemplazados por otros elementos de la población seleccionados al azar.

Para minimizar el riesgo de incurrir en errores producidos por la especificación deficiente de la población, se deben establecer muy claramente los objetivos de la encuesta antes de llevar a cabo el estudio, proporcionando así una imagen clara de los elementos que componen la población. Lo más importante de todo, es que el investigador debe ser muy cuidadoso y referir sus inferencias a la población real de la que se tomó la información muestral y no alguna otra población, quizás conceptual, de mayor interés.

La experiencia es la mejor guía para controlar la fuente de error en las encuestas por muestreo. Los individuos o agencias que han diseñado y efectuado numerosas encuestas de determinado tipo (por ejemplo, de opinión pública, de investigación de mercados, de auditorías contables, de auditorías de inventarios) desarrollan la habilidad de prever determinadas fallas en la encuesta. Por lo anterior, son capaces de diseñar la muestra y los métodos de la encuesta para eliminar las fuentes más comunes de sesgos y error y, al mismo tiempo, minimizar la influencia de las fuentes incontrolables de error.

Nos hemos referido en el primer módulo autoformativo, con relación a que las muestras probabilísticas se caracterizan porque en ellas cada elemento de la población tiene una probabilidad conocida y no nula de ser seleccionado. Dado que la distribución de probabilidades suele ser conocida, en este tipo de muestras es posible proceder a la inferencia estadística, trasladando los datos muestrales a la población. Para ello es necesario hacer las correspondientes estimaciones, en base a los datos muestrales, para proceder, a continuación, al cálculo de su variabilidad, es decir, al cálculo de los errores estándar. Lo anterior no es posible con las muestras no probabilísticas, razón por la cual reafirmamos que nuestro estudio se concentrará en las muestras probabilísticas.

Para que las muestras sean probabilísticas, y, por tanto, reúnan las características señaladas, se han de seguir determinadas normas en el proceso de extracción de la muestra, garantizando la aleatoriedad prevista en el cálculo de probabilidades. Estas normas, ajustadas a las peculiaridades de cada universo o población, dan origen a los cuatro métodos (diseños) básicos de muestreo que cubren distintos procesos de extracción de muestras. Estos métodos son los siguientes:

- Muestreo aleatorio simple.
- Muestreo aleatorio sistemático.
- Muestreo aleatorio estratificado.
- Muestreo aleatorio por conglomerados.

De estos métodos los tres primeros suelen ser monoetápicos. En ellos los elementos de la muestra se eligen directamente en una sola etapa. El último es polietápico, es decir de varias etapas y las unidades muestrales en principio no son los elementos de la población, sino conglomerados (conjunto o grupo) de elementos.

Los cuatro procedimientos de muestreo reseñados son distintos y su elección debe responder a las particularidades del universo o población. Esto quiere decir que, en la mayoría de los casos, no es indiferente utilizar un método u otro, ya que cada uno de ellos suele ser el más indicado para cada situación. No obstante, en la práctica, a veces hay que utilizar no los métodos más precisos sino los más viables, dada la carencia de información sobre el universo o los elevados costos de determinadas aplicaciones. Así, la utilización del muestreo aleatorio simple para la realización de una encuesta nacional

dirigida a la población adulta podría ser, en teoría, la más adecuada, pero; en la práctica, resulta inviable, y lo mismo ocurre en otros muchos casos. Esto quiere decir que, en general, cabe la posibilidad de sustituir unos métodos por otros, mediante la realización de un nuevo diseño y, en especial, la adecuación del tamaño de la muestra al método que se aplique. Esto es así porque, para un error de muestreo dado, varía el tamaño de la muestra en función del tipo de muestreo que se utilice.

Anteriormente hemos señalado que hay cuatro métodos básicos de muestreo. Sin embargo, no se utilizan necesariamente de forma aislada, es mas, muy frecuentemente se aplican simultáneamente en la extracción de una misma muestra. Esta técnica responde a necesidades de diseño y se concreta en la combinación de métodos por etapas o por submuestras. En el primer caso, se puede utilizar un método distinto en cada fase del muestreo y, en el segundo, en cada submuestra.

**Por ejemplo:**

*Para la extracción de una muestra nacional, dirigida a la población en general, se puede proceder, primero, mediante un muestreo por conglomerados para terminar con la elección de las unidades últimas de muestreo, mediante muestreo aleatorio simple. En una muestra a médicos, después de estratificar por especialidades, se puede extraer la muestra, en las especialidades menos numerosas, mediante muestreo aleatorio simple, y, en el resto, después de estratificar nuevamente, bien por municipio/departamento o por lugar donde desarrolla el médico su actividad: hospital, ambulatorio, domicilio.*

En conclusión podemos apreciar que existen varios métodos para diseñar muestras aleatorias, cualquiera de los métodos no está exento de error, debiéndose tener mucho cuidado para que al momento del diseño, desde la planeación se busque como reducir los errores que puedan afectarnos las estimaciones.

*Realizaremos una breve revisión de algunos de los conceptos que hemos estudiado en Estadística relacionado a las estimaciones en el subtema que continúa, esto constituye la base para estudiar los distintos métodos de muestreo y realizar las estimaciones pertinentes.*

### 3. Estimaciones en el muestreo

En el módulo de estadística se ha estudiado que un aspecto de importancia lo constituye el analizar el comportamiento de una variable (característica) para una población de interés. Sin embargo al ser muchas de las poblaciones de interés, demasiado grandes para efectos de poder calcular los parámetros poblacionales hemos planteado que es mucho más razonable pensar en una muestra representativa, a partir de la cual se puedan realizar inferencias sobre muestras tomadas de la población.

Las medidas numéricas usadas para resumir las características de una población de  $N$  elementos son definidas como valores esperados de una variable aleatoria  $x$ . Por

definición, el valor esperado de  $x$ ,  $E(x)$ , está dado por:  $E(x) = \sum_{i=1}^N x_i p(x_i)$ . Recordemos

que teníamos definida la media poblacional por:  $\mu = \frac{\sum_{i=1}^N x_i}{N}$ , podemos observar que:

$E(x) = \sum_{i=1}^N x_i p(x_i) = \frac{1}{N} \left( \sum_{i=1}^N x_i \right) = \mu$ ; esto es que el valor esperado de la variable  $x$  es igual a la media poblacional  $\mu$ , donde  $\mu$  es un parámetro poblacional.

Recordemos la simbología que estamos utilizando:

E: Representa el valor esperado

X: variable aleatoria

N: Número de elementos de la población

$\sum_{i=1}^N$ : Sumatoria de los elementos de la población, desde el 1 hasta el N.

La variabilidad de las mediciones en una población puede ser medida por la varianza, la cual es definida como el valor esperado del cuadrado de la desviación entre una medición  $x$  seleccionada aleatoriamente y su valor medio  $\mu$ . Por lo que la varianza de  $x$ ,  $V(x)$ , está

$$\text{dada por: } V(x) = E(x - \mu)^2 = \sum_{i=1}^N (x - \mu)^2 p(x) = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}.$$

La varianza  $V(x)$  es comúnmente denotada por:  $\sigma^2$  para la población. La desviación estándar de la población, se define como la raíz cuadrada de la varianza poblacional y se denota por:  $\sigma = \sqrt{\sigma^2}$ .

En estudios estadísticos, a la población de interés puede que no les conozcamos las mediciones, es por esa razón que se puede especular acerca de la naturaleza de  $\mu$  y  $\sigma$  por medio de una selección de  $n$  mediciones que denominamos muestra y a partir del estudio de las propiedades de esa muestra, podemos inferir las características de la población siguiendo el mismo patrón establecido para resumir la información de una población.

Podemos calcular entonces la media, la varianza y la desviación estándar de una muestra de tamaño  $n$  por medio de las expresiones:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \text{ donde } \bar{x} \text{ la llamamos media muestral y es un estadístico muestral.}$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \text{ donde } s^2 \text{ la llamamos varianza muestral y es otro estadístico muestral.}$$

$$s = \sqrt{s^2}, \text{ donde le llamamos desviación estándar y es otro estadístico muestral.}$$

Se tratará entonces de utilizar estas cantidades muestrales como aproximaciones razonables para estimar  $\mu$ ,  $\sigma^2$  y  $\sigma$  que son los parámetros poblacionales.

A manera de síntesis, estaremos interesados en calcular estadísticos muestrales para realizar estimaciones de los parámetros poblacionales de una población objeto de estudio.

**Veamos el siguiente ejemplo:**

Liste todas las posibles muestras irrestrictas aleatorias de tamaño  $n = 2$  que pueden seleccionarse de la población  $\{0, 1, 2, 3, 4\}$  y calcule la  $\sigma^2$  de la población.

Las posibles muestras aleatorias de tamaño  $n = 2$  que pueden seleccionarse se pueden establecer en cuanto a cantidad por medio de la expresión que hemos estudiado en estadística y que relaciona el número de combinaciones de los elementos que conformaran las muestras sin que los elementos se repitan, esta expresión está dada por:

$${}_N C_n = \frac{N!}{n!(N-n)!}$$

Para el caso concreto de los datos del ejemplo, se tiene que la combinación nos resulta:

$${}_5 C_2 = \frac{5!}{2!(5-2)!} = \frac{5!}{2! \times 3!} = \frac{5 \times 4 \times 3!}{2! \times 3!} = 10$$

La lista de las muestras que nos resulta son mostradas a continuación donde adicionalmente se presentan después de las flechitas los valores de las medias de cada muestra :

1.  $\{0, 1\} \rightarrow 0.5$
2.  $\{0, 2\} \rightarrow 1$
3.  $\{0, 3\} \rightarrow 1.5$
4.  $\{0, 4\} \rightarrow 2$
5.  $\{1, 2\} \rightarrow 1.5$
6.  $\{1, 3\} \rightarrow 2$
7.  $\{1, 4\} \rightarrow 2.5$
8.  $\{2, 3\} \rightarrow 2.5$
9.  $\{2, 4\} \rightarrow 3$
10.  $\{3, 4\} \rightarrow 3.5$

Usted puede verificar que de la población que se nos suministró, no existe una muestra más de tamaño  $n = 2$ , asimismo para ejercitarnos, comprobemos el valor de las medias que se encuentran para cada una de las diez muestras que se listan.

El valor de la media de la población estará dada por lo siguiente

$$\mu = \frac{4+3+2+1+0}{5} = 2$$

$$\mu = 2$$

Podemos encontrar el valor de la varianza de la población por medio de la expresión que estudiamos en estadística y aplicándolo a los datos suministrados la varianza poblacional resulta:

**Continuando con el ejemplo:**

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

$$\sigma^2 = \frac{(0-2)^2 + (1-2)^2 + (2-2)^2 + (3-2)^2 + (4-2)^2}{5}$$

$$\sigma^2 = \frac{4+1+0+1+4}{5} = \frac{10}{5} = 2$$

Tengamos presente estos mismos cálculos en el segundo tema de la presente unidad autoformativa en donde los retomaremos.

Por otro lado, conviene recordar las características de la distribución probabilística Normal, que nos servirá para la realización de las estimaciones. En la Figura 1 se muestra un bosquejo de una distribución normal de probabilidad:

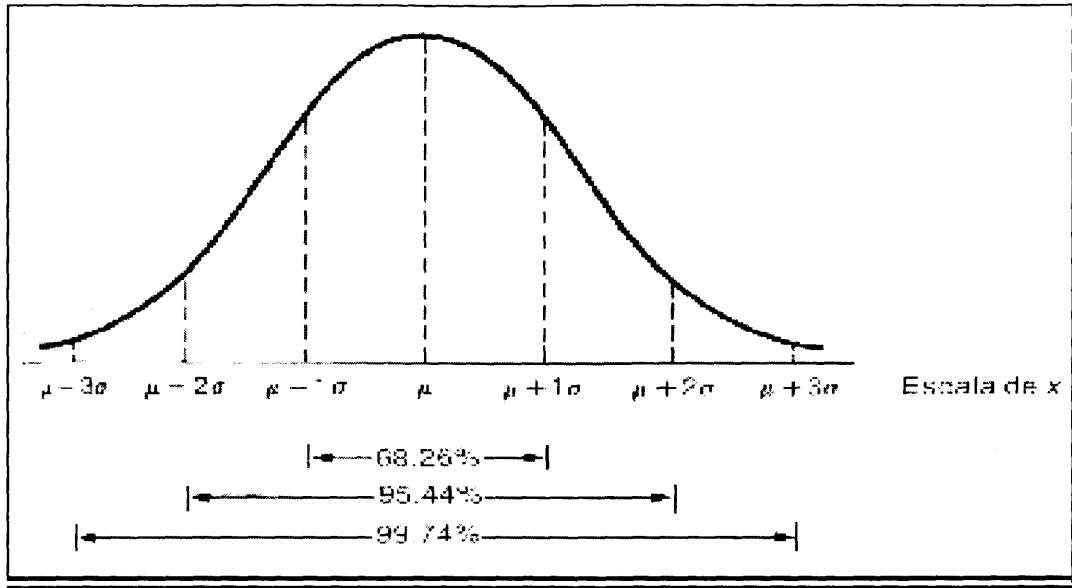


Figura 1

La curva normal tiene perfil de campana y presenta un solo pico en el centro exacto de la distribución.

La media aritmética, la mediana y la moda de la distribución son iguales y están en el punto central.

La mitad del área bajo la curva se encuentra a un lado del valor central  $\mu$  y la otra mitad al otro lado.

La distribución normal es simétrica con respecto a su media.

La curva normal decrece uniformemente en ambas direcciones a partir del valor central  $\mu$ .

Es asintótica, lo cual significa que la curva se acerca cada vez más al eje horizontal, pero en realidad nunca llega a tocarlo.

El modelo matemático de la distribución probabilística normal o función Gaussiana, es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$
, el cual está determinado por dos parámetros poblacionales  $\mu$  y  $\sigma$ , que al ser especificados nos producirá una distribución de probabilidad normal diferente para cada combinación de parámetros.

La distribución probabilística normal estándar, resulta de convertir o estandarizar la distribución normal por medio de una transformación de variable:  $z = \frac{x - \mu}{\sigma}$ .

Al realizar el cambio de variable  $z = \frac{x - \mu}{\sigma}$ , el modelo matemático de la función normal de

probabilidad se estandariza, transformándose en el siguiente modelo:  $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ ,

que se muestra en la Figura 2, donde el eje horizontal es ahora el eje  $z$  y las desviaciones estándar son transformadas a valores numéricos en la escala de  $z$ .

Notar la diferencia para las figuras 1 y 2. En la figura 1,  $x = \mu$  se convierte en  $z = 0$  en la figura 2,  $x = \mu + 1\sigma$  en  $z = 1$ ,  $x = \mu + 2\sigma$  en  $z = 2$ , y así sucesivamente.

Conociendo el valor  $z$ , se puede obtener el área o probabilidad bajo la curva normal estandarizada con la tabla que aparece el anexo 1.

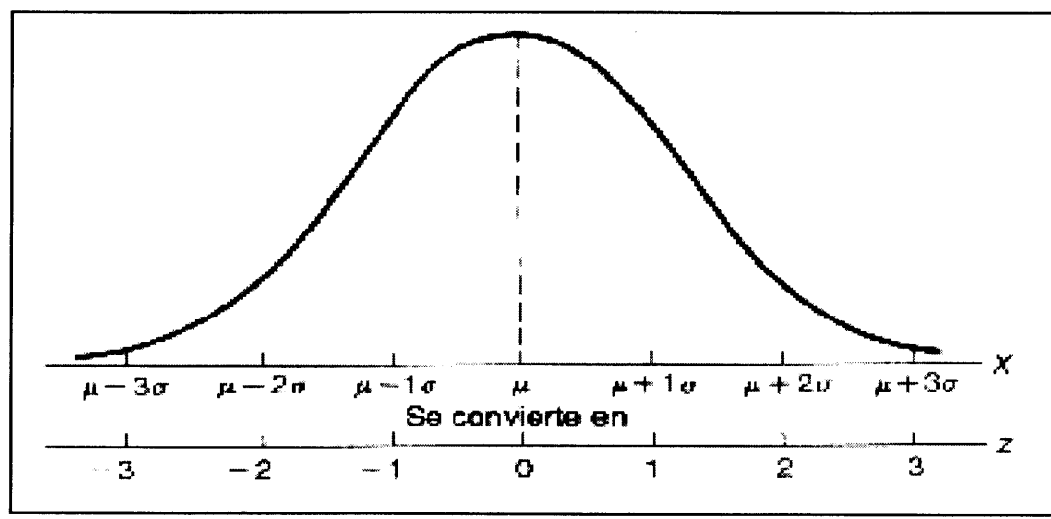


Figura 2.

En la figura 1 también, se muestran distintas áreas bajo la curva normal, de forma sintetizada podemos resumir que:

A una desviación estándar se tiene el 68.25% del área.

A dos desviaciones estándar se tiene el 95.44% del área.

A tres desviaciones estándar se tiene el 99.74% del área.

El área bajo la curva de la distribución normal estandarizada de probabilidad y el eje horizontal, es igual a la unidad (1) y resulta que podemos encontrar el valor de la probabilidad que se le asocia a algún evento por medio del cálculo del área en la distribución que nos muestra la figura 2 y el anexo 1. Recordar que el valor de la probabilidad no puede ser superior a la unidad,  $0 \leq p(x) \leq 1$ .

Esta información nos servirá para poder construir los intervalos de confianza que estudiaremos más adelante en la presente unidad.

Por otro lado retomaremos un teorema fundamental tanto para el desarrollo de la teoría del muestreo como para el de la estimación, este es:

**Teorema:** Sean  $x_1, x_2, \dots, x_n$ ;  $n$  variables aleatorias idénticamente distribuidas con una distribución de probabilidad no especificada y que tienen una media  $\mu$  y una varianza  $\sigma^2$  desconocidas. La media muestral la calcularemos a partir de la siguiente formula:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

La media muestral tiene una distribución con media  $\mu$  y con varianza  $\sigma^2/n$ , que tiende hacia una distribución normal conforme  $n$  tiende a infinito. Dicho de otra manera, la

variable aleatoria  $\frac{(\bar{x} - \mu)}{(\frac{\sigma}{\sqrt{n}})}$  tiene como limite una distribución normal estándar.

Este teorema es el conocido como Teorema del límite central.

Si la muestra es lo suficientemente grande, (esto es si su tamaño es superior a 30 observaciones), sin importar la distribución de la población, la distribución de medias muestrales puede considerarse como una distribución normal. La distribución de medias muestrales está completamente caracterizada por su media  $\mu$  y su desviación estándar a

la que se denomina como error estándar de estimación, en símbolos  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ , debido a que permite cuantificar la desviación de las medias muestrales en relación con el parámetro poblacional.

Observemos de forma intuitiva que el error en la estimación de la media se reduce en la medida en que las muestras son de mayor tamaño, puesto que la desviación estándar de la población esta dividida por la raíz cuadrada del tamaño de la muestra, lo cual concuerda con lo abordado en el subtema anterior relacionado al error de estimación en el muestreo. En el extremo, si el tamaño de la muestra es exactamente igual al tamaño de la población, el error estándar de la estimación será cero.

La teoría del muestreo proporciona las herramientas necesarias para poder hacer inferencias de los valores de parámetros poblacionales desconocidos a partir de una muestra representativa de la población existiendo dos tipos de estimaciones: la estimación puntual y la estimación por intervalo.

### a. Estimación puntual

Estimación puntual se le denomina a un número (estadístico) que se emplea para estimar un parámetro poblacional.

Un estimador puntual es una estimación univaluada (un solo valor) del valor de un parámetro poblacional. La media muestral es un estimador puntual de la media población.

### **Retomemos ejemplo:**

Si en el ejemplo de la edad promedio de los estudiantes de primer ingreso en la carrera de administración de empresas en la UCA, el valor de la edad media que nos resulte de sumar todas las edades obtenidas para los estudiantes que se seleccionaron en la muestra, dividido por el tamaño de la muestra ( $n$ ) y que nos resulte por ejemplo en  $\bar{x} = 18$  años, este valor que es un estadístico muestral, es un estimador puntual que lo emplearemos para estimar el parámetro poblacional, en este caso la edad promedio de la población (estudiantes de primer ingreso en la carrera de administración de empresas en la UCA).

Pero **¿qué es lo que define la calidad en la estimación?**

Es conveniente definir algunos criterios que nos permitirán identificar con claridad los mejores estimadores puntuales:

## **1) Insesgamiento**

Cuando se toma una muestra y se calcula un estadístico muestral para hacer inferencia, se produce un sesgo de estimación que es equivalente a la diferencia entre el valor efectivo de un parámetro poblacional y el valor muestral obtenido. En tal sentido es recomendable que el estadístico muestral tenga un promedio o esperanza matemática igual al parámetro poblacional, esto es, que sea un estimador insesgado del parámetro poblacional.

Si el valor esperado del estadístico muestral es igual al parámetro poblacional que se estima, se dice que dicho estadístico es un estimador insesgado del parámetro

poblacional. Si  $\hat{\theta}$  es un estimador del parámetro  $\theta$ , entonces  $E(\hat{\theta}) = \theta$ .

*Algunos estimadores insesgados.*

- a) La medida de las medias de todas las muestras que pueden tomarse de una población determinada, es exactamente igual a la medida de la media poblacional como lo hemos expresado al inicio del presente subtema.
- b) El estadístico  $s^2$ , es un estimador insesgado de la varianza poblacional  $\sigma^2$ ,

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

puesto que  $E(s^2) = \sigma^2$ , siendo

## **2) Eficiencia**

Recordando que la desviación estándar de las medias muestrales, esto es el error estándar de estimación, es un cuantificador de la exactitud de la estimación. En general,



si existen dos estimadores insesgados de un mismo parámetro poblacional  $\theta$ , se preferirá aquel que presente la menor varianza.

Definición: Sea  $\hat{\theta}$  un estimador de  $\theta$ , tal que  $E(\hat{\theta}) = \theta$ , y la varianza de  $\hat{\theta}$  es menor que la varianza de cualquier otro estimador insesgado de  $\theta$  para todos los posibles valores de  $\theta$ . Se dice entonces que  $\hat{\theta}$  es un estimador insesgado de varianza mínima de  $\theta$ , esto es, un estimador eficiente de  $\theta$ .

### 3) Consistencia

Un estimador puntual es consistente si sus valores tienden a acercarse al parámetro poblacional conforme se incrementa el tamaño de la muestra. Así muestras grandes proporcionarán mejores estimadores puntuales que las muestras pequeñas. La media aritmética de la muestra es un estimador consistente de  $\mu$  puesto que el error estándar de estimación se calcula mediante la relación  $\sigma/\sqrt{n}$ , el que obviamente disminuye en la medida en que aumenta el tamaño de la muestra.

#### b. Estimación por Intervalo

La estimación por intervalo expresa la amplitud dentro de la cual probablemente se encuentra un parámetro poblacional.

El intervalo dentro del que se espera esté un parámetro poblacional, por lo general se denomina intervalo de confianza, si nos fijamos en la figura 1 antes expuesta, podemos identificar lo siguiente:

En 1 caso de 3 (más exactamente el 31.75%), se espera que el área bajo la curva no esté a una desviación estándar, esto es que en 2 casos de 3, el área bajo la curva se encuentre a una desviación estándar esto es el 68.25% de los casos es probable encontrar al parámetro poblacional en el intervalo de confianza del 68.25% que corresponde a una desviación estándar en la distribución normal estandarizada para  $z=1$ .

En 1 caso de 20 (más exactamente el 4.66%), se espera que el área bajo la curva no esté a dos desviación estándar, esto es que en 19 casos de 20, el área bajo la curva se encuentre a dos desviación estándar esto es el 95.44% de los casos es probable encontrar al parámetro poblacional en el intervalo de confianza del 95.44% que corresponde a dos desviaciones estándar en la distribución normal estandarizada para  $z=2$ .

En 1 caso de 100 (más exactamente el 0.26%), se espera que el área bajo la curva no esté a tres desviación estándar, esto es que en 99 casos de 100, el área bajo la curva se encuentre a tres desviación estándar esto es el 99.74% de los casos es probable encontrar al parámetro poblacional en el intervalo de confianza del 99.74% que corresponde a tres desviaciones estándar en la distribución normal estandarizada para  $z=3$ .

**Por ejemplo**

*El intervalo de confianza para la media poblacional es el intervalo que tiene una mayor probabilidad de contener a la media poblacional  $\mu$ . Se utilizan con frecuencia dos intervalos de confianza para la media poblacional: el intervalo de confianza de 95% y el intervalo de confianza de 99%. ¿Pueden utilizarse otros intervalos de confianza?*

**¿Cómo se interpreta, por ejemplo, un intervalo de confianza de 95%?**

Un intervalo de confianza de tal valor significa que aproximadamente 95% de los intervalos contruidos similarmente contendrá el parámetro que se estima. Si se emplea un nivel de confianza de 99%, entonces se esperan que casi el 99% de los intervalos obtenga el parámetro a estimar.

Lo anterior nos muestra dos características importantes:

- 1) No todo intervalo formado incluye al parámetro.
- 2) Si se forman 100 intervalos y se utiliza el nivel de 95%, no exactamente 95 de los intervalos incluirá a dicho parámetro.

Otra interpretación del intervalo de confianza de 95%, es que 95% de las medias muestrales para un tamaño de muestra especificado, estará dentro de 1.96 desviaciones estándar de la media poblacional hipotética. De manera similar, para un intervalo de confianza de 99%, entonces 99% de las medias muestrales estarán dentro de 2.58 desviaciones estándar de la media poblacional hipotética. Notar que los valores 1.96 y 2.58 son los valores de  $z$  y que se han encontrado en la tabla del anexo 1.

**¿Cómo se elabora un intervalo de confianza para la media?**

El procedimiento es el siguiente:

**Primero:** Es necesario calcular el error estándar de la media, se calcula por medio de:

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ . En este caso se supone conocida la desviación estándar de la población.

Si no se conoce y  $n \geq 30$ , la desviación estándar de la media muestral, denotada por  $s$ , sirve para aproximar la desviación estándar de la población,  $\sigma$ . Entonces la formula para

el error estándar quedará como:  $S_{\bar{x}} = \frac{S}{\sqrt{n}}$ .

Observemos que el error estándar de la media variará de acuerdo con el tamaño de la muestra que está en el denominador. A medida que aumenta cada vez más el tamaño de la muestra,  $n$ , la variabilidad de las medias se vuelve cada vez más pequeña.

Lógicamente, una estimación de la media poblacional basada en una muestra grande es más confiable que una estimación realizada con una muestra pequeña.

**Segundo:** Elaboración de los intervalos de confianza de 95% y de 99%. Estos intervalos de confianza se establecen como sigue cuando  $n \geq 30$ .

Intervalo de confianza de 95%:  $\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$

Intervalo de confianza de 99%:  $\bar{x} \pm 2.58 \frac{s}{\sqrt{n}}$

En general:

Intervalo de confianza:  $\bar{x} \pm z_{\alpha} \frac{s}{\sqrt{n}}$

Donde  $z_{\alpha}$  es un valor denominado coeficiente de confianza

$\frac{s}{\sqrt{n}}$  es el error estándar de la media,

$\alpha$  es un valor que se le denomina nivel de significación y que tiene relación con el error de tipo I que se ha estudiado en estadística.

$\bar{x}$  es la media muestral.

Los valores extremos  $\bar{x} - z_{\alpha} \frac{s}{\sqrt{n}}$  y  $\bar{x} + z_{\alpha} \frac{s}{\sqrt{n}}$  del intervalo de confianza se denominan límites de confianza. La medida de confianza que se obtiene se denomina grado o nivel de confianza, esto es 0.95, 0.99.

Observemos que al tener un 99% de confianza, el intervalo de confianza es mayor que al tener un 95% de confianza, teniéndose por lo tanto una pérdida de precisión en la estimación ya que se aumenta el intervalo de confianza.

### Ejemplo:

Si en nuestro ejemplo de la edad promedio de los estudiantes de primer ingreso en la carrera de administración de empresas en la UCA, se ha obtenido de la muestra un valor de  $\bar{x} = 18$  años, el valor de la desviación estándar muestral  $s$  que se obtiene al calcular la desviación estándar de los valores obtenidos por todas las mediciones de  $x_i$  en la muestra nos resulta  $s = 6$  años y el tamaño de la muestra ha sido  $n = 36$ , se obtendrá:

Intervalo de confianza al 95%:  $18 \pm 1.96 \frac{6}{\sqrt{36}} = 18 \pm 1.96 \rightarrow (16.04 ; 19.96)$  lo que significa que el

95% de las medias muestrales para un tamaño de  $n = 36$  estará dentro de 1.96 desviaciones estándar de la media poblacional hipotética, esto es, la edad promedio de la población constituida por los estudiantes de primer ingreso en la carrera de administración de empresas en la UCA, se encuentra entre los 16.04 y 19.96 años con un nivel de confianza del 95%.

Para el caso de un intervalo de confianza del 99%, se tiene:

$18 \pm 2.58 \frac{6}{\sqrt{36}} = 18 \pm 2.58 \rightarrow (15.44 ; 20.58)$  lo que significa que el 99% de las medias muestrales

para un tamaño de  $n = 36$  estará dentro de 2.58 desviaciones estándar de la media poblacional. Observar que el intervalo de confianza para el 99% es mayor que para el 95%, lo que representa una disminución de la precisión en la estimación, al aumentarse el intervalo de confianza.

Generalizando: Para tener mayores niveles de confianza en la estimación, es necesario perder precisión, aumentando el intervalo de confianza o viceversa. La única manera de aumentar la precisión sin modificar el nivel de confianza es aumentando el tamaño de la muestra, disminuyendo de esta manera el error estándar de estimación.

### Ejemplo:

Observemos que para disminuir el intervalo de confianza para el 99% en el ejemplo antes visto, manteniéndose la desviación estándar de la muestra, es conveniente incrementar el tamaño de la muestra, con ello se mantiene la precisión (si lo comparamos con el 95% de confianza, al compararlo con el 99%, la precisión se incrementa) al disminuirse el error estándar de estimación. Es así que al incrementar el tamaño de la muestra a  $n = 64$  por ejemplo, tendremos para un intervalo de confianza del 99%:  $18 \pm 2.58 \frac{6}{\sqrt{64}} = 18 \pm 1.935 \rightarrow (16.065 ; 19.935)$ .

De forma similar cuando deseemos estimar la proporción poblacional desconocida  $p$ , siendo  $\bar{p}$  la proporción muestral; la estimación puntual para la proporción de población es:

$$\bar{p} = \frac{\text{Número de Éxitos en la Muestra}}{\text{Número Muestreado}} = \frac{x}{n}$$

El intervalo de confianza para una proporción de la población es:

$$\bar{p} \pm z\sigma_{\bar{p}}$$

donde  $\sigma_{\bar{p}}$  es el error estándar de la proporción:  $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$

Por lo tanto, el intervalo de confianza se establece mediante:  $\bar{P} \pm Z \sqrt{\frac{P(1-P)}{n}}$

Cuando las poblaciones se consideran muy grandes o se supone que son infinitas, es aplicable el conjunto de expresiones anteriormente presentadas para el error estándar y para los intervalos de confianza tanto para la media como para la proporción. Cuando la población muestreada no es infinita, ni siquiera es muy grande (en esto se ha omitido los cálculos para nuestro ejemplo de los estudiantes de primer ingreso), se debe de realizar ajustes en el error estándar de la media y en el error estándar de la proporción por medio de:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$

Este ajuste se denomina factor de corrección por población finita.

### Ejemplo:

En el caso de nuestro ejemplo es una población finita, dado que tiene un límite superior fijo del tamaño de la población y que lo hemos asumido en  $N = 385$ .

En base al factor de corrección por población finita ( $\sqrt{\frac{N-n}{N-1}}$ ) para el cálculo del error

estándar de la media  $\sigma_x = \frac{\sigma}{\sqrt{n}}$ , recalcular tanto los errores estándar así como los intervalos de confianza para la media, considerando que nuestra población es finita.

Si la muestra es un porcentaje considerable de la población, se esperaría que cualesquiera estimaciones fueran más precisas que para muestras más pequeñas. La regla general es que si la proporción  $n/N$  es de menos de 0.05, se omite el factor de corrección por población finita dado que el impacto del factor de corrección es bastante pequeño.

### Me ejercito:

Para la misma muestra de tamaño  $n=36$  de los estudiantes de primer año de Administración de Empresas de la UCA, se ha obtenido que 20 son mujeres. Encuentre el intervalo de confianza del 95% para la proporción de mujeres de los estudiantes de primer año que estudian Administración de Empresas en la UCA.

### c. Solución del tamaño de la Muestra:

Hasta ahora, se han especificado los tamaños de muestra pero es conveniente tener una expresión que nos permita determinar el tamaño adecuado de muestra.

Se debe tener cuidado de no seleccionar una muestra demasiado grande o demasiada pequeña. Si se selecciona una muestra demasiado grande, se estará derrochando inútilmente tiempo y dinero. Si se selecciona una muestra demasiado pequeña, las conclusiones a las que se llegará acerca de la población podrían ser incorrectas.

Hay varios errores comunes acerca del tamaño adecuado de una muestra, algunos de estos son:

- 1) Que un porcentaje constante semejante respecto a la población sea adecuado para todo problema.
- 2) Una muestra más grande debe seleccionarse para un municipio con alta densidad poblacional como Masaya que para un municipio de baja densidad poblacional como Condega.

Los elementos anteriores son realmente falsos dado que los factores que determinan el tamaño de la muestra, ninguno tiene que ver directamente con el tamaño de la población. Los tres factores que determinan el tamaño de la muestra son:

- 3) El grado de confianza seleccionados. Por lo general es de 0.95 o 0.99, puede ser cualquier otro nivel. Es el investigador el que determina el grado de confianza. Ya hemos visto que cuanto más alto sea el grado de confianza, tanto mayor será la muestra necesaria para tener cierta precisión.

- 4) El máximo error permisible. El investigador debe también decidirlo. Es el máximo error tolerable en un nivel de confianza específico.
- 5) La variación de la población. La variación o variabilidad de la población la mide la desviación estándar o la varianza.

Observemos que  $s_x = \frac{s}{\sqrt{n}}$  es el error estándar de la media.

Pero ese error estándar lo podemos considerar adicionalmente como el error total permisible relacionado con el coeficiente de confianza que hemos asignado, esto es:

$$s_x = \frac{\text{Error Total Permissible} \rightarrow B}{z \text{ Desviación Estándar} \rightarrow z}$$

$s_x = \frac{B}{z} = \frac{s}{\sqrt{n}}$ , donde  $s$  es la desviación estándar de la muestra,  $n$  es el tamaño de la muestra,  $B$  es el error total permisible y  $z$  es el grado o coeficiente de confianza seleccionado.

A partir de la igualdad anterior, se despeja el valor  $n$ , encontrándose que

$$n = \left( \frac{z \times s}{B} \right)^2$$

Donde:

$B$ : Error total permisible.

$z$ : es el desvío normal asociado al grado de confianza seleccionado.

$s$ : es la desviación estándar de la muestra.

Como  $s$  no es conocida porque se quiere determinar el tamaño de la muestra, convendrá estimar la variación de la población por medio de:

- Realizar una encuesta piloto y utilizar la desviación estándar.
- Calcular la desviación estándar con base en un conocimiento de la población.

Como podemos observar, el tamaño de la muestra que se calcula usando la fórmula no siempre es el correcto dado que es poco probable que la desviación estándar de la muestra o del estudio piloto sea el correcto; pero al menos nos brinda un valor para calcular el tamaño de la muestra.

En el caso del tamaño de muestra para proporciones, se deben especificar tres aspectos:

- El investigador debe decidir que nivel de confianza emplear, por lo general de 0.95 o 0.99.
- Debe indicar que tan precisa debe ser la estimación de la proporción de la población.
- La proporción de la población,  $\bar{p}$ , se debe aproximar con base en la experiencia.
- La formula para determinar el tamaño de la muestra en el caso de una proporción

es: 
$$n = \bar{p}(1 - \bar{p}) \left( \frac{z}{B} \right)^2$$

donde:

$\bar{p}$  : es la proporción estimada, con base en la experiencia o en un estudio piloto.

Z = es el desvío normal z asociado al grado de confianza seleccionado.

B: es el error máximo permisible que el investigador tolerará.

Notemos que estas expresiones para el calculo del tamaño de la muestra se consideran para una población muy grande o infinita, lo cual no involucra la inclusión del factor de corrección para población finita.

Tener presente que lo abordado en este primer tema de la presente unidad autoformativa nos brinda los elementos para delimitar adecuadamente los pasos 3, 4 y 8 de la planeación de una encuesta por muestreo, estudiada en el tema D, de la primera unidad autoformativa.

El paso 3 corresponde al establecimiento del marco muestral, este deberá desprenderse de la población objeto de estudio, habiéndose identificado claramente las unidades muestrales y de ahí estructurar el listado que constituye el marco muestral.

El paso 4 corresponde al diseño de muestreo, aquí se ha definido lo que es un diseño de muestreo, teniéndose presente que en los dos temas siguientes de la presente unidad autoformativa y en la tercera, se estudiará cada uno de los cuatro procedimientos de muestreo probabilístico.

En cuanto al número de elementos de la muestra (tamaño de la muestra), se ha establecido que el investigador deberá tener presente tres factores para su cálculo, estos son:

- El grado de confianza, este factor es determinado por el investigador, recomendándose el 95% de nivel de confianza.
- El máximo error permisible, esto es el máximo error (B) que el investigador está dispuesto a tolerar, este valor es determinado por el investigador. Tenga presente que el error estará asociado a la variable que estemos estudiando y al estimador que utilicemos para realizar las inferencias. Es así que como en el ejemplo que hemos estado estudiando, si el estimador es una media de la edad, el valor de B deberá ser razonable y en las dimensiones del estimador, esto es en años. Si el estimador es una proporción, el error deberá ser un porcentaje de la proporción. Si el estimador es un total poblacional, el error deberá ser correspondiente al estimador.

Para cada tipo de procedimiento de muestreo que se estudiará, se relacionan fórmulas para el cálculo del tamaño de la muestra, debiendo seleccionarse el valor de B por su parte.

- La variabilidad de la población. Este factor es desconocido, por cuanto lo que se desea con el muestreo es que a partir de una muestra representativa se realicen inferencias sobre la población. Como se ha mencionado en el presente tema, debemos realizar una prueba piloto y calcular la desviación estándar de los datos obtenidos en la prueba piloto para la variable que se estudia, esta desviación estándar será sustituida por la desviación estándar de la población.

El paso 8, corresponde a la prueba piloto, esta como se ha establecido en su oportunidad es una pequeña muestra que se selecciona de la población o marco muestral y que nos permite comprobar varios elementos importantes como lo son:

- El instrumento de medición (tema D de la primera unidad autoformativa, relacionada al diseño de formas para la recolección de datos)
- Calificar a los entrevistadores
- Verificar el manejo de las operaciones de campo
- Calcular la desviación estándar de los datos obtenidos sobre la variable más importante que hemos seleccionado para sustituirla por la desviación estándar poblacional y proceder al cálculo del tamaño de la muestra. La prueba piloto deberá ser de un tamaño de entre 30 a 50 elementos de la población o marco muestral para que tenga cierto nivel de confianza el cálculo de la desviación estándar que se sustituirá por la desviación estándar poblacional en las fórmulas correspondientes para el cálculo del tamaño de la muestra.

Con los conceptos y elementos hasta ahora recordados y estudiados, tenemos los elementos necesarios y suficientes para desarrollar nuestro estudio del primer método de muestreo que estudiaremos: el muestreo aleatorio simple, así como también todos los elementos para completar el plan para el trabajo final.

### Actividad de Autoaprendizaje No. 1

1. Establezco el objetivo que se persigue al realizar una encuesta por muestreo de una población bajo estudio.
2. Con mis propias palabras defino lo que entiendo por error de estimación.
3. Si el parámetro poblacional posee un valor de 12 unidades y el estimador muestral tiene un valor de 11.5, ¿cuál es el valor del error de estimación? ¿Qué es entonces el valor de B?
4. ¿Qué factores debo elegir para el cálculo del tamaño de la muestra?. Explico!
5. Una muestra de 49 observaciones se toma de una población normal. La media muestral es de 55, y la desviación estándar de la muestra es 10. Determino el intervalo de confianza de 99% para la media de la población.
6. Asumo que uno de los canales de Televisión de Nicaragua, planea sustituir uno de sus programas que se transmite en el horario con mayor número de televidentes con un nuevo programa educativo orientado a la familia. Antes que se tome la decisión, se tomó una muestra aleatoria de 400 personas que acostumbran presenciar programas en el horario citado. Después de ver una edición de prueba del programa educativo, 250 de las personas consultadas indicaron que sí verían el nuevo programa.
  - a) ¿Cuál es la estimación de la proporción de Televidentes en la población que verá el nuevo programa educativo?
  - b) Definir un intervalo de confianza de 95% para la proporción de personas que verá el nuevo programa.
  - c) La estimación de la proporción de la población ha de estar dentro de  $\pm 0.05$ , con un nivel de confianza de 95%. La mejor estimación de la proporción poblacional es 0.15. ¿De qué tamaño se requiere una muestra?

Busco al final de la unidad autoformativa II, las respuestas de esta actividad de autoaprendizaje, en la página No. 182 y me retroalimento.



## B. MUESTREO ALEATORIO SIMPLE

### 1. Definición

Al seleccionar una muestra de  $n$  mediciones de una población finita de  $N$  mediciones, si el muestreo se lleva a cabo de forma que todas las muestras posibles de tamaño  $n$  tengan la misma probabilidad de ser seleccionadas, el muestreo se llama aleatorio y el resultado es una muestra aleatoria simple.

Recordemos nuestro ejemplo del subtema C de la presente unidad autoformativa donde la población que teníamos era constituida por los elementos  $\{0,1,2,3,4\}$ , ahora bien, las muestras de tamaño  $n = 2$  que podemos formar con la población de  $N = 5$ , nos resultó  ${}_N C_n = 10$  en el ejemplo, por lo que si estuviésemos interesados en elegir de esa población una sola muestra de tamaño  $n = 2$ , la probabilidad de elegir una muestra determinada por ejemplo la  $\{0,1\}$ , es de  $1/10$ . Observemos que conocemos la probabilidad de elegir una muestra en particular, esto sólo es posible en los muestreos de carácter probabilísticos que ya hemos mencionado en la unidad anterior.

Sin embargo el procedimiento elegido anteriormente resulta ser tedioso cuando los tamaños de la población y de la muestra se incrementan, por ejemplo si estuviésemos interesados en extraer una muestra de tamaño  $n = 10$  en una población de  $N = 50$ , el número de posibles muestras será:

$${}_N C_n = \frac{50!}{10! \times (50-10)!} = \frac{50!}{10! \times 40!} = \frac{50 \times 49 \times 48 \times 47 \times 46 \times 45 \times 44 \times 43 \times 42 \times 41 \times 40!}{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \times 40!} = 10,272,278,170$$

Significaría que deberíamos generar más de diez mil millones de muestras de tamaño  $n = 10$  de esa población de tamaño  $N = 50$ , lo cual nos haría inoperante el procedimiento de selección de una muestra.

Dada esta situación, para elegir una muestra de tamaño  $n = 10$  en una población de tamaño  $N = 50$ , tendremos que elegir al azar a diez miembros de la población que serán los que conformaran la muestra de tamaño  $n = 10$ , en vez de detallar cada una de las muestras de tamaño  $n = 10$  que se pueden obtener de la población de  $N = 50$ .

En este tipo de muestreo entonces, la selección de los elementos de la muestra se hace en una sola etapa, directamente y sin reemplazamiento. Por las razones antes expuestas es que se tiene que elegir de forma directa de la población a cada elemento de la muestra y sin regresar (reemplazamiento) al elemento que ya fue elegido de la población y que forma parte de la muestra para que no vuelva a aparecer dado que si esto sucede no nos brindará más información, en otras palabras no nos conviene obtener por dos veces la misma información dado que ya no será relevante para nuestros propósitos.

En la práctica equivale a censar o utilizar el censo de la población objeto del estudio, para sacar, después, al azar, los elementos que van a formar parte de la muestra.

**Así por ejemplo:**

*Si se quiere extraer una muestra de 150 elementos, de un universo formado por los médicos en activo en Nicaragua, habría, en primer lugar, que obtener el censo de los 12000 médicos, que aproximadamente existen, y, a partir de aquí, se haría la selección de los individuos. Esta se puede hacer mediante tablas de números aleatorios u otro procedimiento similar, como la tómbola, que también garantiza la aleatoriedad.*

Las tablas de números aleatorios simplifican considerablemente el proceso. Presentan múltiples combinaciones de números extraídos al azar y, con ellas, a partir de cualquier fila o columna, se toman tantos números consecutivos, como el de elementos de la muestra. Los números sacados al azar indican, sobre un universo numerado, los elementos que se incorporan a la muestra y que, por tanto, deben ser entrevistados.

El muestreo aleatorio simple se aplica, fundamentalmente, en investigaciones sobre poblaciones pequeñas, plenamente identificables, máxime si constituyen universos específicos y diferenciados. Así una muestra sobre médicos o sobre cualquier otro colectivo diferenciado y pequeño se debería extraer mediante muestreo aleatorio simple. Es más, en muchas ocasiones hay que acudir a este tipo de muestreo empezando por la creación de un censo del universo objeto de estudio.

En el caso de poblaciones grandes, ofrece mayores dificultades. En primer lugar, es difícil realizar un listado de todo el universo; si ello fuera posible, el proceso para extraer la muestra resultaría tedioso; finalmente, la dispersión de la muestra podría ser muy importante, lo que implicaría aumento considerable de los costos y del tiempo del trabajo de campo. Así, la realización de una muestra de 1200 elementos dirigida a la población Nicaragüense en edad de trabajar, y utilizando el muestreo aleatorio simple, sería impensable. En primer lugar habría que reconstruir y hacer el listado del universo, con mas de un millón de unidades. Después, habría que seleccionar, aleatoriamente, los elementos de la muestra para, finalmente, proceder al trabajo de campo, es decir, a la recogida de la información. La primera dificultad radicaría en la reconstrucción del universo, operación prácticamente imposible. Aunque se dispusiera del censo, la dispersión de las entrevistas sería muy grande. Junto a una cierta concentración en las grandes ciudades, se produciría la dispersión en pequeños pueblos, valles y comarcas en los lugares más recónditos. La muestra podría ser, técnicamente, perfecta, pero los costos y el tiempo empleado serían mucho mas elevados que por otros procedimientos. De ahí que, cuando hay que hacer muestras sobre poblaciones muy grandes, se recurra a otros métodos de muestreo o a la combinación de varios.

El muestreo aleatorio simple se presenta como el prototipo de muestreo por su sencillez y la facilidad para calcular los errores de muestreo. Esto obedece a que es un muestreo monoetápico, las diferentes unidades del universo tienen la misma probabilidad de formar parte de la muestra.

## **2. Procedimiento de selección de una muestra aleatoria simple**

**Primero:** Se requiere realizar previamente una numeración de 1 a N de todos los individuos de la población en estudio.

### **A manera de ejemplo:**

Supongamos que tenemos una población de interés de tamaño  $N=50$ , constituida por un grupo de estudio modalidad regular de Técnicas de Recopilación de la Información y deseamos elegir una muestra de  $n=10$  estudiantes.

El primer paso nos exige que debemos de numerar de 1 a 50, todos los miembros de la población, es así que a cada uno de los estudiantes del grupo se le asigna un número del 1 al 50, sin repetirse, uno a uno, donde a todos los miembros de la población debe asignársele un número y cada número del 1 al 50 debe ser asociado a cada elemento de la población.

**Segundo:** Generamos  $n$  números aleatorios de entre los 1 a  $N$  números que representan a los individuos de la población. La generación de los números, la realizamos mediante una tabla de números aleatorios, por medio de una tómbola o de una computadora.

### **Continuando con el ejemplo anterior:**

Como la muestra de nuestro interés es de tamaño  $n = 10$ , entonces debemos elegir diez números aleatorios que estén en el rango de 1 a 50, esto es porque si nos resultan por ejemplo números aleatorios que no estén en dicho rango, no tendremos la forma de identificar cuál será nuestro elemento de la muestra dado que no existen elementos con numeraciones mayores a 50 en el ejemplo que estamos siguiendo.

Este segundo paso nos plantea que la generación de los números aleatorios se realice por cualquiera de los tres siguientes procedimientos: La Tómbola, la tabla de números aleatorios y un programa de computadora, por ejemplo el Excel.

### **Ejemplo:**

La tómbola es el procedimiento que aplica la Lotería Nacional para sus sorteos ordinarios y extraordinarios, consiste en un cilindro cerrado que contiene unas pelotillas numeradas que se hace girar varias veces y que tiene un punto(hueco) único por donde se extrae una a una las pelotillas numeradas (en este caso se pueden regresar las pelotillas a la tómbola) y se va anotando el número que va saliendo hasta completar los cinco dígitos de los números de la Lotería que participan en cada sorteo. El rango de los números va de 1 a 40,000.

La tabla de números aleatorios, es una tabla de números que se ha construido a partir de un algoritmo matemático conteniendo lo que se le denomina una raíz de aleatorización. El procedimiento que se sigue es cerrar los ojos y con el dedo o lápiz después de varios movimientos ubicar un número en la tabla donde cayó la punta del dedo o del lápiz.

### **Ejemplo:**

Si el primer número, por ejemplo fue el 68766 y a partir de ahí procedemos hacia abajo, o hacia arriba o hacia los lados a ir siguiendo el número siguiente hasta poder completar el tamaño de muestra que se requiere. Por ejemplo seguiremos hacia abajo del número que se eligió inicialmente, obteniendo los números siguientes:

1) 68766    2) 88159    3) 97054    4) 53746    5) 54194    6) 24374    7) 15181    8) 82845  
9) 52498    10) 57179    11) 93634    12) 69903    13) 97728    14) y así sucesivamente.

### **Para la elección de los miembros de la muestra, ¿cómo lo hacemos?**

#### **Siguiendo con nuestro ejemplo:**

El rango está de 1 a 50 para nuestra población, entonces debemos elegir números aleatorios que estén en ese rango, es así que como los números aleatorios son de cinco dígitos, tomamos solamente dos dígitos de cada número (dado que los números de 1 a 50 solamente tienen dos dígitos) pueden ser los dos primeros o los dos últimos; para nuestro ejemplo tomemos los dos últimos pero iremos descartando aquellos números que excedan el valor de 50, esto es, tomaremos de los dos últimos dígitos aquellos números que se encuentren en el rango, el resto los descartamos.

Es así que el primer número de acuerdo a los criterios de elección es el 66, entonces lo descartamos; el siguiente el 59 que se descarta, el tercero es el 54 que se descarta, el cuarto es el 46 que no se descarta y así sucesivamente.

De los trece números que hemos representado en el material cinco son números que concuerdan con los criterios y se encuentran en el rango de 1 a 50, estos son:

El cuarto y corresponde al número 46

El octavo y corresponde al número 45

El onceavo y corresponde al número 34

El doceavo y corresponde al número 03

El treceavo y corresponde al número 28

**Ejercítase** en la elección de los restantes números aleatorios y complete los diez números que debe tener nuestra muestra.

Tener presente que si se repite el número que cumple las condiciones se deberá seguir el procedimiento descartando el número que salió por segunda vez, hasta completar todos los números que el tamaño de la muestra nos exige, en nuestro ejemplo  $n = 10$ .

Por medio del programa Excel, este es otro procedimiento en donde podemos aprovechar la hoja de cálculo Excel. Ubicados en una celda cualquiera, aplicamos función (fx) en el icono superior derecho y se nos desplegará un cuadro con dos opciones generales: Categoría de la Función y Nombre de la Función. En categoría de la función, seleccionamos Matemáticas y trigonométricas y en nombre de la función, buscamos la función "ALEATORIO. ENTRE" y le damos aceptar, esta función nos pide un número inferior y un número superior. El número inferior corresponde al valor menor en el rango, en nuestro ejemplo es el valor de 1; el número superior corresponde al valor mayor de nuestro rango, para nuestro ejemplo corresponde al número 50. El rango que hemos definido para nuestro ejemplo es el (1,50). Al darle aceptar, nos generará un número aleatorio en dicho rango y después copiamos en 9 o 10 celdas para generar 10, 11 números aleatorios que cumplen la condición de estar en el rango. Si los números se repiten, favor descartarlos, debemos elegir solamente 10 números aleatorios en nuestro caso. Al ser aplicado por ejemplo nos resultó una serie con los números aleatorios siguientes: 34, 45, 31, 29, 47, 30, 31, 25, 9, 49 y 18. De todos ellos descartamos una vez el número repetido que es el 31, quedándonos los diez números aleatorios.

Observar que si usted realiza varias veces el presente procedimiento en dicho rango, los resultados que obtendrá serán distintos. **¿Porqué?**

**Tercero:** Elegir los miembros de la muestra de tamaño  $n$  por medio de los números generados al azar, haciendo coincidir los  $n$  números generados con los números correspondiente de la población. Este tercer paso nos permite realizar una asociación entre los " $n$ " números aleatorios generados.

### **Continuación del ejemplo:**

En nuestro caso  $n = 10$  con los números del 1 a  $N$  de nuestra población, en nuestro ejemplo 1 a 50, es así que los miembros de la población que van a formar parte de la muestra serán:

<u>Número Aleatorio:</u>	<u>Número en la Población:</u>	<u>Miembro de la población elegido para la muestra:</u>
46	46	Juan Pérez
45	45	Luz Marina Araica
34	34	Carlos Solís
03	03	Maria Barrios
28	28	Sandra Ordóñez

Los otros cinco miembros de la muestra han sido seleccionados por usted!

### Ejercítese:

Favor asuma que su grupo de clase es una población a ser estudiada y de ella debe elegir una muestra de 8 elementos. Aplique completamente el procedimiento de elección de una muestra aleatoria simple basándose en la tabla de números aleatorios o en el programa Excel.

Prepárese a compartir en la sesión tutorial con el resto del grupo de clase el procedimiento y la muestra elegida.

Ya teniendo elegidos todos los miembros que forman parte de la muestra, se procede al levantamiento de los datos de interés que se han establecido para el desarrollo del estudio. Esto se realiza por medio de la aplicación del instrumento de medición que se ha diseñado, obteniéndose el suministro de los datos de parte de todos los miembros de la población que han sido elegidos como miembros de la muestra.

### 3. Estimaciones basada en una muestra aleatoria simple

Una vez recolectadas las observaciones muestrales, el siguiente objetivo consiste en la estimación de ciertos parámetros poblacionales de interés. Frecuentemente se tiene interés en estimar la media poblacional ( $\mu$ ), la proporción poblacional ( $p$ ) o el total poblacional  $\tau$  (letra griega tau).

La media poblacional, ya se ha definido en estadística como el valor de una característica de la población y que representamos por:  $\mu = \frac{\sum x_i}{N}$ , esto es por ejemplo el salario medio de los trabajadores públicos de Managua, que nos resulta de tomar los salarios de todos los trabajadores públicos de Managua, sumarlos y dividirlo por el número de trabajadores.

La proporción poblacional también se ha definido en estadística como:  $p = \frac{X}{N}$ , siendo  $X$  los elementos de la población que poseen un atributo particular de interés y  $N$  el número de elementos de la población, es así que la proporción poblacional se define como la fracción, porción relativa o porcentaje que expresa la parte de la población que tiene ese atributo particular de interés. Si el atributo de interés por ejemplo es  $X = \text{"sexo femenino"}$  en los trabajadores públicos de Managua, entonces  $p = \frac{X}{N}$  es la proporción de trabajadores públicos mujeres de Managua, siendo  $N$  el total de todos los trabajadores públicos de Managua.

El total poblacional es el valor que resulta de sumar de todos los miembros de la población la característica que se estudia, es así que si sumamos por ejemplo todos los salarios del mes de octubre de los trabajadores públicos de Managua,  $\sum x_i$ , obtendremos el total poblacional, que resulta ser:  $\tau = \sum x_i = \mu N$ , esto es la media poblacional ( $\mu$ ) multiplicada por el número de elementos que existen en la población ( $N$ ). Es así que por ejemplo, una firma de contabilidad podría estar interesada en el valor promedio de las cuentas por cobrar y también en el valor total de estas cuentas, así como en la proporción de clientes que tienen cuentas por cobrar.

Las fórmulas de cálculo para estimar la media poblacional  $\mu$ , el total poblacional  $\tau$ , y la proporción poblacional  $p$ , basados en un muestreo aleatorio simple, no pueden quedarse como una simple estimación puntual,  $\bar{\mu}$ ,  $\bar{\tau}$  y  $\bar{p}$ , dado que no proporcionan información

acerca de la bondad de la estimación. Por lo que es conveniente incluir las fórmulas de la varianza con el propósito de establecer cotas para el error de estimación de los parámetros  $\mu$ ,  $\tau$  y  $p$ , y asegurar el calculo para el límite del error de estimación, realizando por lo tanto estimaciones por intervalo.

#### 4. Estimador de la media poblacional $\mu$ , varianza estimada de la media y límite para el error de estimación de la media

El estimador de la media poblacional  $\mu$ , como ya lo hemos establecido anteriormente se

calculará por medio de la siguiente fórmula:  $\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$  fórmula 2.1

La Varianza estimada de la media  $\bar{x}$ , se calculará por medio de:  $v(\bar{x}) = \frac{s^2}{n} \left( \frac{N-n}{N} \right)$ , fórmula 2.2

$$\text{donde: } s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}$$

El Limite para el error de estimación, lo calcularemos multiplicando el valor de  $z$  para el nivel de confianza elegido por el investigador, por la raíz cuadrada de la varianza

estimada de la media  $\bar{x}$ , esto es:  $z\sqrt{\hat{v}(\bar{x})} = z\sqrt{\frac{s^2}{n} \left( \frac{N-n}{N} \right)}$ , fórmula 2.3

donde  $\frac{N-n}{N}$  es un factor de corrección por población finita.

Observar que cuando el valor de  $N$  es muy, muy grande o infinito, el factor de corrección por población finita  $\frac{N-n}{N}$  tiende hacia 1. Entonces la varianza estimada de la media  $\bar{x}$

se calcula por medio de  $V(\bar{x}) = \frac{s^2}{n}$  y el limite para el error de estimación de la media  $\bar{x}$

se calcula por medio de  $z\sqrt{\frac{s^2}{n}}$ .

**Veamos los siguientes ejercicios para consolidar las fórmulas antes planteadas:**

Liste todas las posibles muestras irrestrictas aleatorias de tamaño  $n = 2$  que pueden seleccionarse de la población  $\{0,1,2,3,4\}$ . Esta población ya la conocemos!. Calcule la  $\sigma^2$  de la población y la  $V(\bar{x})$  de la media muestral y luego, demuestre por cálculos directos que

$$V(\bar{x}) = \frac{N-n}{N-1} \left( \frac{\sigma^2}{n} \right)$$

**Solución:** Recordemos que esa población es la misma que se nos había solicitado anteriormente que encontraríamos las muestras de tamaño  $n=2$  y la varianza de la población. Para repasar, las muestras de tamaño 2 que se nos generan de la población son diez y que se determinan por la expresión de las combinaciones siguiente:

$${}_N C_n = \frac{N!}{n!(N-n)!}$$

Resolviendo se tiene:

$${}_N C_n = \frac{5!}{2!(5-2)!} = \frac{5!}{2! \times 3!} = \frac{5 \times 4 \times 3!}{2! \times 3!} = 10$$

Las diez muestras que se generan son:

$$\begin{array}{lllll} 1. \{0, 1\} \rightarrow 0.5 & 2. \{0, 2\} \rightarrow 1 & 3. \{0, 3\} \rightarrow 1.5 & 4. \{0, 4\} \rightarrow 2 & 5. \{1, 2\} \rightarrow 1.5 \\ 6. \{1, 3\} \rightarrow 2 & 7. \{1, 4\} \rightarrow 2.5 & 8. \{2, 3\} \rightarrow 2.5 & 9. \{2, 4\} \rightarrow 3 & 10. \{3, 4\} \rightarrow 3.5 \end{array}$$

Tener presente que las muestras no deben de repetirse, Hemos ubicado al lado de cada muestra generada los valores que tienen la media para cada muestra.

A nivel de la población se calcula la media y la Varianza, obteniéndose lo siguiente:

$$\begin{aligned} \mu &= \frac{4 + 3 + 2 + 1 + 0}{5} = 2 & \mu &= 2 \\ \sigma^2 &= \frac{\sum (x_i - \mu)^2}{N} \\ \sigma^2 &= \frac{(0-2)^2 + (1-2)^2 + (2-2)^2 + (3-2)^2 + (4-2)^2}{5} \\ \sigma^2 &= \frac{4+1+0+1+4}{5} = \frac{10}{5} = 2 \end{aligned}$$

La varianza de las medias de las muestras  $V(\bar{x})$  se determina por el valor esperado que ya hemos estudiado, teniendo por lo tanto:

$$\begin{aligned} V(\bar{X}) &= E(\bar{x} - \mu)^2 = \sum (\bar{x} - \mu)^2 p(x) \\ &= (0.5-2)^2 \left(\frac{1}{10}\right) + (1-2)^2 \left(\frac{1}{10}\right) + (1.5-2)^2 \left(\frac{2}{10}\right) + (2-2)^2 \left(\frac{2}{10}\right) + (2.5-2)^2 \left(\frac{2}{10}\right) + (3-2)^2 \left(\frac{1}{10}\right) + (3.5-2)^2 \left(\frac{1}{10}\right) \\ &= \frac{2.25+1+0.5+0+0.5+1+2.25}{10} = \frac{7.50}{10} = 0.75 \end{aligned}$$

Luego con base a la expresión que se nos ha suministrado en el ejercicio se tiene:

$$V(\bar{X}) = \frac{N-n}{N-1} \left( \frac{\sigma^2}{n} \right) = \frac{5-2}{5-1} \left( \frac{2}{2} \right) = \frac{3}{4} = 0.75$$

demostrándose por lo tanto lo que se nos había solicitado.

**Veamos otro ejercicio:**

Para las muestras irrestrictas aleatorias generadas en el ejercicio anterior, calcule la  $s^2$  de cada muestra. Demuestre numéricamente que

$$E(s^2) = \frac{N}{N-1} \sigma^2$$

**Solución:**

Recordemos que:  $E(x_i) = \sum_{i=1}^n x_i P(x_i)$  y  $s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1}$

Para las muestras que hemos generado en el ejercicio anterior, debemos encontrar las varianzas muestrales para cada muestra, obteniéndose los siguientes resultados:

$$1 - s^2 = \frac{(0 - 0.5)^2}{1} + \frac{(1 - 0.5)^2}{1} = (0.5)^2 + (0.5)^2 = 0.25 + 0.25 = 0.5$$

$$2 - s^2 = \frac{(0 - 1)^2}{1} + \frac{(2 - 1)^2}{1} = 1 + 1 = 2$$

$$3 - s^2 = \frac{(0 - 1.5)^2}{1} + \frac{(3 - 1.5)^2}{1} = 2.25 + 2.25 = 4.5$$

$$4 - s^2 = \frac{(0 - 2)^2}{1} + \frac{(4 - 2)^2}{1} = 4 + 4 = 8$$

$$5 - s^2 = \frac{(1 - 1.5)^2}{1} + \frac{(2 - 1.5)^2}{1} = 0.25 + 0.25 = 0.5$$

$$6 - s^2 = \frac{(1 - 2)^2}{1} + \frac{(3 - 2)^2}{1} = 1 + 1 = 2$$

$$7 - s^2 = \frac{(1 - 2.5)^2}{1} + \frac{(4 - 2.5)^2}{1} = 2.25 + 2.25 = 4.5$$

$$8 - s^2 = \frac{(2 - 2.5)^2}{1} + \frac{(3 - 2.5)^2}{1} = 0.25 + 0.25 = 0.5$$

$$9 - s^2 = \frac{(2 - 3)^2}{1} + \frac{(4 - 3)^2}{1} = 1 + 1 = 2$$

$$10 - s^2 = \frac{(3 - 3.5)^2}{1} + \frac{(4 - 3.5)^2}{1} = 0.25 + 0.25 = 0.5$$

El valor esperado de las varianzas muestrales resulta ser :

$$E(s^2) = \sum s_i^2 P(s_i^2) = (0.5)\left(\frac{4}{10}\right) + (2)\left(\frac{3}{10}\right) + (4.5)\left(\frac{2}{10}\right) + (8)\left(\frac{1}{10}\right) = 0.2 + 0.6 + 0.9 + 0.8 = 2.5$$

$$E(s^2) = 2.5$$

En base a la expresión que se nos ha suministrado, se tiene que :

$$\frac{N}{N-1} \sigma^2 = \frac{5}{5-1} (2) = \frac{10}{4} = 2.5 \Rightarrow E(s^2) = \frac{N}{N-1} \sigma^2$$

donde se demuestra lo que se nos ha solicitado.

Favor verificar el calculo de las medias y varianzas tanto para este ejercicio como para el anterior.



**Me ejercito:**

Siguiendo los pasos de los dos ejercicios anteriores y con los mismos datos de la población:  $\{0, 1, 2, 3, 4\}$ , liste todas las posibles muestras aleatorias de tamaño  $n = 3$  que pueden seleccionarse de la población. Calcule la  $\sigma^2$  de la población y la  $V(\bar{x})$  de la media muestral y luego, demuestre por cálculos directos que

$$V(\bar{x}) = \frac{N-n}{N-1} \left( \frac{\sigma^2}{n} \right)$$

calcule la  $s^2$  de cada muestra y demuestre numéricamente que

$$E(s^2) = \frac{N}{N-1} \sigma^2$$

¿Qué sucede con el valor del error para las muestras, cuando  $n$  crece?

Prepárese para discutirlo y socializarlo en clase.

**Veamos otro ejercicio:**

Un psicólogo desea estimar el tiempo de reacción promedio para un estímulo entre 100 pacientes de un hospital especializado en trastornos nerviosos. Una muestra irrestricta aleatoria de  $n = 20$  pacientes fue seleccionada, y fueron medidos sus tiempos de reacción, con los resultados siguientes  $\bar{x} = 2.1$  segundos  $s = 0.4$  segundos. Estime la media poblacional  $\mu$ , y establezca un límite para el error de estimación.

Los datos que tenemos son:

$$N = 100$$

$$n = 20$$

$$\bar{x} = 2.1 \text{ segundos}$$

$$S = 0.4 \text{ segundos}$$

$$\mu = ?$$

La Varianza estimada de  $\bar{x}$  se encuentra con la fórmula 2.2 :

$$V(\bar{x}) = \frac{S^2}{n} \left( \frac{N-n}{n} \right) = \frac{(0.4^2)}{20} \left( \frac{100-20}{100} \right) = 0.0064$$

Límite para el error de estimación a 95% de nivel de confianza se encuentra con 2.3 :

$$Z\sqrt{V(\bar{x})} = 1.96\sqrt{0.0064} = 0.1568$$

$$\mu \rightarrow \bar{x} \pm Z\sqrt{V(\bar{x})} = 2.1 \pm 0.1568$$

$$\mu \rightarrow (1.9432, 2.2568)$$

De donde se tiene que la media poblacional de tiempos medidos de reacción para los 100 pacientes del hospital especializado se encuentra en un 95% de nivel de confianza entre 1.9432 y 2.2568 segundos.

Observar que  $1.9432 = 2.1 - 0.1568$  y corresponde al límite inferior del intervalo;  $2.2568 = 2.1 + 0.1568$  y corresponde al límite superior del intervalo.

**Me ejercito:**

Si la población de pacientes es de 300 y la muestra que se toma es de 45,  $\bar{x} = 2.0$  segundos y  $s = 0.5$  segundos, estime nuevamente el tiempo de reacción promedio y establezca un límite para el error de estimación con el 99% de confianza. Me preparo para socializarlo con el resto de mis compañeros.

## 5. Estimador del total poblacional $\tau$ , varianza estimada del total y límite para el error de estimación del total poblacional

El estimador del total poblacional  $\tau$ , se calcula por medio de lo siguiente:

$$\hat{\tau} = N\bar{x} = \frac{N \sum_{i=1}^n x_i}{n}, \text{ fórmula 2.4}$$

Varianza estimada de  $\tau$  es calculada por:

$$\hat{v}(\hat{\tau}) = \hat{v}(N\bar{x}) = N^2 \left( \frac{s^2}{n} \right) \left( \frac{N-n}{N} \right), \text{ fórmula 2.5}$$

donde

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

El Límite para el error de estimación del total poblacional se calcula por:

$$z\sqrt{\hat{v}(N\bar{x})} = z\sqrt{N^2 \left( \frac{s^2}{n} \right) \left( \frac{N-n}{N} \right)}, \text{ fórmula 2.6}$$

**Ejercitación:**

¿Qué parecido y/o que diferencia aprecia entre las fórmulas de varianzas estimadas para  $\bar{x}$  y  $\tau$ ? Favor prepárese para socializarlo en la sesión tutorial.

**Veamos un ejercicio para reforzar la utilización de las fórmulas:**

Un auditor detecta que cierta empresa regularmente exagera las cantidades de dinero en córdobas de los inventarios, a causa de los retrasos en el registro de los retiros. El auditor quiere estimar la cantidad total exagerada sobre 1,000 artículos listados, obteniendo las cantidades exactas (revisadas) del inventario mediante una muestra aleatoria de 15 artículos, y comparando esas cifras exactas con las cifras registradas. Los datos para los artículos seleccionados se muestran en la siguiente tabla con valores en córdobas. Estime la cantidad total exagerada sobre los 1,000 tipos de artículos, y establezca un límite para el error de estimación. (Omita el factor de corrección por población finita en los cálculos)

Artículo número	Cantidad revisada	Cantidad registrada	Exageración (Diferencia)
1	175	210	35
2	295	305	10
3	68	91	23
4	74	82	8
5	128	140	12
6	241	250	9
7	362	384	22
8	72	80	8
9	59	82	23
10	112	140	28
11	118	124	6
12	210	230	20
13	240	260	20
14	123	247	24
15	96	108	12

### Solución:

Aplicando la fórmula 2.4, se encuentra el estimador del total poblacional  $\tau$ , para ello calculamos la media y el total poblacional.

$$\bar{x} = \frac{\sum \text{Diferencias}}{n} = \frac{260}{15} = 17.33333$$

$$\bar{\tau} = N\bar{x} = 1000 * (17.33333) = 17,3333.33$$

Al omitirse el factor de corrección por población finita, la varianza estimada de  $\bar{\tau}$  es calculada por la fórmula 2.5 modificada, que es:

$$V(\bar{\tau}) = V(N\bar{x}) = N^2 \left( \frac{s^2}{n} \right)$$

donde  $s^2 = 75.24$ . Este valor se encuentra al calcular las varianzas de las diferencias.

Favor verifiquelo!

$$V(\bar{\tau}) = 1000^2 * \frac{75.24}{15} = 5,016,000$$

El límite para el error de estimación es calculado por la fórmula 2.6

$$z \sqrt{N^2 \left( \frac{s^2}{n} \right)} = 4,389.70$$

Por lo tanto:

$$\tau \rightarrow \bar{\tau} \pm z \sqrt{V(\bar{\tau})} = 17,333.33 \pm 4,389.70$$

$$\tau \rightarrow (12,543.63; 21,723.03)$$

Por lo que la cantidad total exagerada sobre los 1,000 tipos de artículos se encuentra entre C\$ 12,543.63 y C\$ 21,723.03 a un nivel de confianza del 95%.

**Veamos otro ejercicio:**

El auditor de una empresa esta interesado en estimar el numero total de comprobantes de viáticos que fueron archivados incorrectamente. En una muestra irrestricta aleatoria de  $n = 50$  comprobantes tomados de un grupo de  $N = 250$ , 20 fueron archivados incorrectamente. Estime el numero total de comprobantes archivados incorrectamente de los  $N = 250$ , y establezca un limite para el error de estimación. [Sugerencia: Si  $p$  es la proporción poblacional de comprobantes archivados incorrectamente, entonces  $N\bar{p}$  es el numero total de comprobantes archivados incorrectamente. Un estimador de  $Np$  es  $N\bar{p}$ , el cual tiene una varianza estimada dada por  $N^2V(\bar{p})$ .]

Los datos que tenemos son:  $n = 50$        $N = 250$        $x = 20$        $\bar{p} = \frac{20}{50} = 0.4$

El número total de comprobantes archivados incorrectamente es  $N\bar{p}$ :

$$\hat{\tau} = N\bar{p} = 250 (0.4) = 100$$

varianza estimada de  $\bar{p}$ :

$$\hat{v}(\bar{p}) = \frac{\bar{p}\bar{q}}{n-1} \left( \frac{N-n}{N} \right) \text{ donde } \bar{q} = 1 - \bar{p}, \bar{q} = 0.6$$

$$\hat{v}(\bar{p}) = \frac{(0.4)(0.6)}{50-1} \left( \frac{250-50}{250} \right) = \frac{48}{12,250} = 0.00391837$$

En base a la sugerencia

$$\hat{v}(\hat{\tau}) = N^2 \hat{v}(\bar{p}) = (250)^2 (0.00391837) = 244.898125$$

Por lo que a un nivel de confianza del 95%, el límite para el error de estimación está dado por:

$$Z\sqrt{\hat{v}(\hat{\tau})} = 1.96\sqrt{244.898125} = 30.67247361$$

$$\tau \rightarrow \hat{\tau} \pm Z\sqrt{\hat{v}(\hat{\tau})} = 100 \pm 30.6724$$

$$\tau = (69, 131)$$

La estimación del total poblacional de comprobantes de viáticos archivados incorrectamente que le resulta al auditor de la empresa, se encuentran entre 69 y 131 a un nivel de confianza del 95%.

## 6. Estimador de la proporción poblacional $p$ , varianza estimada de la proporción y límite para el error de estimación de la proporción poblacional

El estimador de la proporción poblacional  $p$  se encuentra por medio de:

$$\hat{p} = \bar{p} = \frac{x}{n} : \frac{\text{Número de éxitos en la muestra}}{\text{Número muestreado}}, \text{ fórmula 2.7}$$

La Varianza estimada de  $\hat{p}$  se calcula por:  $\hat{v}(\hat{p}) = \frac{\hat{p}\hat{q}}{n-1} \left( \frac{N-n}{N} \right)$ , fórmula 2.8

donde:  $\hat{q} = 1 - \hat{p}$

El Límite para el error de estimación se encuentra por medio de:

$$z \sqrt{\hat{v}(\hat{p})} = z \sqrt{\frac{\hat{p}\hat{q}}{n-1} \left( \frac{N-n}{N} \right)}, \text{ fórmula 2.9}$$

**Veamos el siguiente ejercicio para aplicar las fórmulas:**

Una muestra irrestricta aleatoria de  $n = 40$  estudiantes de un colegio fue entrevistada para determinar la proporción de estudiantes que esta a favor del cambio del sistema semestral al trimestral. Veinticinco de los estudiantes respondieron afirmativamente. Estime la proporción de estudiantes del colegio que esta a favor del cambio (suponga que  $N = 2000$ ). Establezca un límite para el error de estimación.

Los datos que se nos suministra son:  $n = 40$      $x = N^\circ$  de éxitos = 25     $N = 2,000$

$\bar{p}$ , debemos estimarlo y establecer un límite para el error de estimación

$\bar{p}$  lo calculamos aplicando la fórmula 2.7

$$\bar{p} = \frac{x}{n} = \frac{25}{40} = 0.625$$

La varianza estimada de  $\bar{p}$  se calcula por medio de la fórmula 2.8:

$$\hat{v}(\bar{p}) = \frac{\bar{p}\bar{q}}{n-1} \left( \frac{N-n}{N} \right) \text{ donde } \bar{q} = 1 - \bar{p} \cdot \bar{q} = 0.375$$

$$\hat{v}(\bar{p}) = \frac{(0.625)(0.375)}{40-1} \left( \frac{2,000-40}{2,000} \right) = \frac{459.375}{78,000} = 0.005889$$

El Límite para el error de estimación a un nivel de confianza del 95% lo calculamos por la fórmula 2.9:

$$Z \sqrt{\hat{v}(\bar{p})} = 1.96 \sqrt{0.005889} = 0.1504$$

$$p \rightarrow \bar{p} \pm Z \sqrt{\hat{v}(\bar{p})} = 0.625 \pm 0.1504$$

$$p \rightarrow (0.4746 \cdot 0.7754)$$

En conclusión, se estima que entre el 47.46% y el 77.54% de la población de estudiantes está a favor del cambio del sistema semestral al trimestral, con un nivel de confianza del 95%.

## 7. Cálculo del tamaño de la muestra para la media poblacional, el total poblacional y la proporción poblacional

### a. Tamaño de la muestra para la media poblacional ( $\mu$ )

El Tamaño de muestra requerido para estimar  $\mu$  con un límite para el error de estimación B se calcula por medio de la siguiente fórmula:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}, \text{ fórmula 2.10}$$

donde  $D = \frac{B^2}{z^2}$

$\sigma^2$  : es la varianza poblacional, que al no conocerse debemos aplicar una prueba piloto para encontrar la varianza de la muestra piloto ( $s^2$ ) y sustituirla por la varianza poblacional.

B: Es el error máximo permisible y es establecido por el investigador, recordar que debe estar dado en las dimensiones del estimador y debe ser pertinente con dicho estimador.

Z: Es el grado de confianza, este factor es elegido por el investigador. A un nivel de confianza del 95%,  $z = 1.96$ .

### ***Veamos el siguiente ejercicio:***

*En el ejercicio resuelto del psicólogo que deseaba estimar el tiempo de reacción promedio para un estímulo, ¿de qué tamaño deberá ser la muestra para estimar  $\mu$  con un límite para el error de estimación de 1 segundo? Use 1.0 segundos como una aproximación de la desviación estándar poblacional.*

Los datos que nos suministran son:

$n$ , es desconocido

$B = 1$  segundo

$\sigma = 1.0$  segundo

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \text{ donde } D = \frac{B^2}{Z^2}$$

con  $Z = 1.96$  a un nivel de confianza del 95%

$$D = \frac{1^2}{1.96^2} = 0.2603$$

$$n = \frac{250(1.0)^2}{(250-1)(0.2603) + (1.0)^2} = \frac{250}{65.8167} = 3.79$$

$n = 4$

*Según nuestros cálculos, el tamaño de la muestra debe ser de  $n = 4$  para estimar el promedio del tiempo de reacción a los estímulos con un error de estimación de 1 segundo y utilizando una aproximación de 1.0 segundos en la desviación estándar poblacional.*

Observar que para este ejercicio, se nos ha suministrado la varianza poblacional, esto en la práctica comúnmente no ocurre, debiéndose realizar la prueba piloto y sustituir la varianza poblacional por la varianza encontrada en la prueba piloto como lo hemos mencionado anteriormente.

### **b. Tamaño de la muestra para estimar el total poblacional ( $\tau$ )**

El Tamaño de muestra requerido para estimar  $\tau$  con un límite para el error de estimación B se encuentra por medio de:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}, \text{ fórmula 2.11}$$

$$\text{donde } D = \frac{B^2}{Z^2 N^2}$$

**Veamos el siguiente ejercicio:**

En el ejercicio del auditor que deseaba conocer la cantidad total exagerada de los 1,000 tipos de artículos, encontrar el tamaño de la muestra para estimar  $\tau$  con  $B = C\$ 1,000$ .

**Solución:**

El cálculo de  $n$  se realiza con la fórmula 2.11, para ello encontramos el valor de  $D$ :

$$D = \frac{B^2}{z^2 N^2} = \frac{1000^2}{(1.96)^2 * (1000)^2} = 0.2603$$

$$S^2 = 75.24 \quad \text{donde}$$

$$n = \frac{1000 * (75.24)}{(999) * (0.2603) + 75.24} = \frac{75240}{335.2879} = 224.4$$

Para realizar la estimación de  $\tau$  con  $B = C\$ 1,000$  y un nivel de confianza del 95%, se requiere un tamaño de muestra de  $n = 225$  artículos.

**c. El tamaño de la muestra para estimar la proporción poblacional ( $\rho$ )**

El Tamaño de muestra requerido para estimar  $p$  con un límite para el error de estimación  $B$  se encuentra por medio de:

$$n = \frac{Npq}{(N-1)D + pq}, \text{ fórmula 2.12}$$

$$\text{donde } q = 1 - p \text{ y } D = \frac{B^2}{Z^2}$$

**Veamos el siguiente ejercicio:**

¿De qué tamaño debe ser la muestra de estudiantes del colegio que les ha consultado el cambio del sistema semestral al trimestral para estimar  $p$  con un error de estimación de  $B = 0.08$ ?

**Solución:**

Aplicando la fórmula 2.12 y con base en los datos suministrados, se tiene que:

$$N = 2000, \quad p = 0.625, \quad q = 0.375, \quad z = 1.96, \quad B = 0.08, \quad D = 0.00166597$$

$$n = \frac{2000 * 0.625 * 0.375}{(2000 - 1) * 0.00166597 + 0.625 * 0.375} = 131.5$$

Se requiere por lo tanto una muestra de tamaño  $n = 132$  para realizar la estimación de  $p$  con un error del 8% a un nivel de confianza del 95%.

A partir de las formulas estudiadas para el muestreo aleatorio simple, solucione la siguiente actividad de autoaprendizaje.

## Actividad de autoaprendizaje No. 2

1. Con los mismos datos de la población:  $\{0,1,2,3,4\}$ , elaboro una lista de todas las posibles muestras aleatorias de tamaño con  $n = 3$  que pueden seleccionarse de la población. Calculo la  $\sigma^2$  de la población y la  $V(\bar{x})$  de la media muestral y luego, demuestre por cálculos directos que

$$V(\bar{x}) = \frac{N-n}{N-1} \left( \frac{\sigma^2}{n} \right)$$

calculo la  $s^2$  de cada muestra y demuestro numéricamente que

$$E(s^2) = \frac{N}{N-1} \sigma^2$$

¿Explico lo qué sucede con el valor del error para las muestras, cuando  $n$  crece?

2. Una muestra irrestricta aleatoria de  $n = 100$  medidores de agua es controlada dentro de una comunidad para estimar el promedio de consumo de agua diario por casa, durante un periodo estacional seco. La media y la varianza muestrales fueron  $\bar{x} = 12.5$  y  $s^2 = 1252$ . Si suponemos que hay  $N = 10,000$  casas dentro de la comunidad, estimo  $\mu$ , el promedio de consumo diario verdadero, y establezco un límite para el error de estimación.
3. Dos dentistas A y B, hicieron una encuesta para investigar el estado de los dientes de 200 niños en una colonia. El Dr. A seleccionó una muestra aleatoria simple de 20 niños y contó el número de dientes cariados de cada niño, con los siguientes resultados:

Número de dientes cariados por niño	0	1	2	3	4	5	6	7	8	9	10
Número de niños	8	4	2	2	1	1	0	0	0	1	1

- a. El Dr. B usando las mismas técnicas dentales examinó a los 200 niños y sólo registró aquellos que no tenían los dientes cariados. Encontró que 60 niños no tenían dientes dañados.
- b. Estimo el número total de dientes cariados en los niños de la colonia, usando:
- c. Sólo los resultados de A
- d. Usando los resultados de A y B
- e. ¿Son insesgadas las estimaciones?
- f. Qué estimación espera que sea más precisa?
4. En una muestra aleatoria simple de 200 obtenida de una población de 2000 colegios, 120 de éstos estuvieron a favor de una propuesta, 57 se opusieron y 23 se abstuvieron de opinar. Estimo los límites de confianza al 95% para el número de colegios en la población que favorecieron la propuesta.
5. En un estudio sociológico, realizado en una pequeña ciudad para estimar la proporción de hogares donde habita cuando menos una persona mayor de 65 años de



edad; se tomó una muestra aleatoria de 60 hogares de los 621 que existen. Al terminar la investigación de campo, de los 60 hogares muestreados, en 11 habitan cuando menos una persona mayor de 65 años de edad. Estimo la proporción poblacional  $p$ , y establezco un límite para el error de estimación.

6. En el ejercicio anterior, ¿de qué tamaño deberá ser la muestra para estimar  $p$  con un límite para el error de estimación de 0.08?. Supongo que la proporción verdadera  $p$  es aproximadamente 0.2
7. El gerente de un taller de maquinaria desea estimar el tiempo promedio que necesita un operador para terminar una tarea sencilla. El taller tiene 98 operadores. Se seleccionaron aleatoriamente a ocho operadores y se les tomó el tiempo. Los resultados obtenidos se muestran en la tabla adjunta. Estimo el tiempo promedio para terminar la tarea entre todos los operadores y establezco un límite para el error de estimación

Tiempo (en minutos)	
4.2	5.3
5.1	4.6
7.9	5.1
3.8	4.1

8. Elaboro el diseño de muestreo, incluyendo la prueba piloto con base a lo aprendido en los primeros dos temas estudiados en la presente unidad autoformativa y las Orientaciones para la elaboración del trabajo final presentada en la primera unidad autoformativa y con base a la Matriz del Estudio que fue elaborada en esa unidad autoformativa.

Encontraré las respuestas a esta actividad de autoaprendizaje en las páginas 182 y 183, al final de la unidad autoformativa II.

## C. MUESTREO ALEATORIO SISTEMÁTICO

### 1. Definición

Una muestra obtenida al seleccionar aleatoriamente un elemento de los primeros  $k$  elementos en el marco y después cada  $k$ -ésimo elemento se denomina muestra sistemática de 1 en  $k$ .

#### ¿Cómo se selecciona $k$ en una situación dada?

Si el tamaño de la población  $N$  es conocido y al determinar un tamaño de muestra  $n$ , se selecciona  $k$  para obtener el tamaño de muestra,  $k$  debe ser menor o igual que  $N/n$

$$(k \leq \frac{N}{n}).$$

#### Veamos el ejemplo:

En nuestro ejemplo de los estudiantes de Técnicas de recopilación de la información, donde  $N = 50$  y se desea seleccionar una muestra de tamaño  $n = 10$  por muestreo aleatorio sistemático, el valor

de  $K$  nos resulta de dividir  $N$  entre  $n$ , esto es  $K = \frac{50}{10} = 5$ . Para este caso  $K \leq \frac{N}{n}$ , cumpliendo

la condición.

Si se tuviese que seleccionar una muestra de tamaño  $n = 12$ , por ejemplo, el valor de

$\frac{N}{n} = \frac{50}{12} = 4.16$ , entonces el valor de  $K$  que se elige es  $K = 4$  para cumplir la condición. Notar

que  $K$  debe ser un número entero y que debe de cumplir la condición de  $K \leq \frac{N}{n}$ .  $K$  es denominado coeficiente de elevación y es un equivalente de subrango en que hay que dividir la población para elegir la muestra de tamaño  $n$ .

En nuestro ejemplo original, si  $K = 5$ ,  $n = 10$  observemos que nuestra población  $N = 50$  es como que la subdividimos en subrangos de la forma siguiente:

1-5, 6-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50 pudiendo notarse que la totalidad de subrangos corresponde a  $n = 10$ .

### 2. Procedimiento de selección de una muestra aleatoria sistemática

**Primero:** Seleccionar  $k$  (coeficiente de elevación), esto se hace al dividir el número de miembros de la población ( $N$ ) entre el número de miembros de la muestra ( $n$ ) que necesitamos.  $K = N/n$ , debiéndose seleccionar el valor de  $k$  entero, que cumpla la condición  $k \leq N/n$ .

**Segundo:** Una vez seleccionado  $k$ , se elige al azar un número entre 1 y  $k$  (primer subrango) por medio de las tablas de números aleatorios u otro procedimiento. Este

número elegido es el punto de arranque de la selección. En el ejemplo que hemos seguido como  $K=5$ , elegiremos un número al azar entre 1 y 5 ( $K=5$ ) por medio de un procedimiento de elección aleatoria nos resultó por ejemplo el número 2 que se encuentra en el subrango 1-5. Si nos resulta un número al azar que no este en el subrango, continuaremos con el procedimiento hasta encontrar el número al azar que se encuentre en el subrango, notar que perfectamente nos pueden resultar los números extremos del subrango.

**Tercero:** Una vez elegido el número al azar, se le suma a éste de forma sucesiva el coeficiente de elevación, dando lugar con ello a los números que pasan a formar parte de la muestra. La elección de la muestra entonces se realiza a partir del número al azar encontrado en el paso anterior, para nuestro ejemplo es el 2 y sumándole sucesivamente el valor que encontramos para el coeficiente de elevación, en nuestro caso  $K=5$ , por lo que la muestra se formará de los números: 2, 7, 12, 17, 22, 27, 32, 37, 42, 47.

Notar que la población la debíamos inicialmente numerar del 1 al 50. Los números que nos resultaron de la elección de la muestra sistemática tienen por decirlo así un nombre y un apellido en la población.

Se puede considerar este método como una variante del método aleatorio simple, en la que difiere la forma de selección de las unidades muestrales. Para ello, se halla el coeficiente de elevación:  $N/n$  y se elige al azar un número no superior al mismo, que es el que indica el punto de arranque de la selección. A partir de aquí, al número elegido se le suma, sucesivamente, el coeficiente de elevación, dando lugar a los números que pasan a formar parte de la muestra. En la encuesta a médicos que hemos utilizado anteriormente como ejemplo, el proceso sería el siguiente: coeficiente de elevación  $N/n = 12000/150 = 80$ ; primer número al azar, no superior a 80, el 24, por ejemplo. La selección sería 24,  $24+80=104$ ,  $104+80=184$ ; y así sucesivamente.

Con este procedimiento se simplifica considerablemente la selección, pero existe el riesgo de introducir sesgos en la muestra al elegir los elementos de forma periódica. Esto ocurre cuando el universo está ordenado en función de determinados criterios que pueden inducir a que la selección sistemática recaiga en elementos que no son representativos de la heterogeneidad del universo. En la selección sistemática de una muestra sobre listas de 20 individuos en que los 10 primeros fueran varones y los 10 últimos mujeres, si el coeficiente de elevación fuera 20, siempre saldrían sólo hombres o sólo mujeres, pero no habría posibilidades de que en la muestra estuvieran representados ambos sexos.

### **a. Ventajas de la selección sistemática**

Entre alguna de las ventajas de la selección sistemática, podemos mencionar las siguientes:

- 1) Tiene las ventajas de extender la muestra sobre toda la población.
- 2) Ser de fácil aplicación, y
- 3) Conseguir un efecto similar al de la estratificación si las unidades se han ordenado previamente siguiendo un cierto criterio como, por ejemplo, colocando al principio de la lista las unidades densamente pobladas y continuando progresivamente hasta llegar a las menos pobladas.

## b. Desventajas de la selección sistemática

Por el contrario, como inconvenientes puede introducir sesgos debido al hecho de que cada unidad en la muestra es seleccionada con una periodicidad constante igual a  $K$ .

Notar que la precisión de las estimaciones por muestreo sistemático depende del orden de las unidades de muestreo en el marco.

Para evitar este peligro hay que hacer un estudio previo del universo de los listados que lo contienen, incluso, si fuera necesario, "desordenarlos" para preparar la selección de los elementos muestrales. Una técnica utilizada, con frecuencia, cuando se realizan muestreos sistemáticos a partir de ficheros de individuos, almacenados en soporte magnético consiste en romper el orden actual del fichero: orden alfabético, de antigüedad, etcétera, a partir de cualquier criterio como, por ejemplo utilizando, las 7 últimas cifras de la cédula de identidad de los individuos que componen el universo, y ordenándolas al revés. Así, el número de identidad 001110360 se quedaría en 1110360 y ordenado al revés sería 0630111. Posteriormente, se procedería al ordenamiento de todo el universo y a la aplicación del muestreo sistemático.

El muestro sistemático es aplicable en el control de calidad industrial, a manera de ejemplo un plan de inspección para los artículos fabricados a lo largo de una línea de montaje en movimiento puede requerir la inspección de cada artículo 50 de la línea. Asimismo los investigadores de mercado y los encuestadores, quienes muestrean personas en movimiento, muy frecuentemente emplean un diseño sistemático. A cada décimo cliente en un mostrador de pago se le puede preguntar su opinión acerca del sabor, color o textura del producto alimenticio.

Como hemos mencionado anteriormente con relación a la precisión de las estimaciones en el muestreo sistemático, estas dependen del ordenamiento de la población, a continuación diferenciaremos algunos conceptos de población de acuerdo con el criterio del ordenamiento.

## 3. Definición de poblaciones aleatorias, ordenada, periódica y su efecto en la varianza y las estimaciones

La varianza de  $\bar{x}$  (en el muestreo aleatorio simple) esta dada por:

$$V(\bar{x}) = \frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right),$$
 mientras la varianza  $\bar{x}_s$  (en el muestreo aleatorio sistemático) está dada por:

$$V(\bar{x}_s) = \frac{\sigma^2}{n} [1 + (n-1)\rho],$$
 donde  $\rho$  es una medida de la correlación entre los pares de elementos dentro de la misma muestra sistemática. Si  $\rho$  está cercano a uno, entonces los elementos dentro de la muestra son bastante similares con respecto a la característica que se esta midiendo, y el muestreo sistemático producirá una varianza de

la media muestral mayor que la obtenida con el muestreo aleatorio simple. Esto es

$$V(\bar{x}_s) = \frac{\sigma^2}{n} (1 + (n-1)\rho) = \sigma^2 > \frac{\sigma^2}{n} = V(\bar{x}).$$
 Si  $\rho$  es negativo, entonces el muestreo

sistemático puede ser mejor que el muestreo aleatorio simple. La correlación puede ser negativa si los elementos dentro de la muestra sistemática tienden a ser extremadamente diferentes.

Para  $\rho$  cercano a cero y  $N$  bastante grande, el muestreo sistemático es aproximadamente equivalente al muestreo aleatorio simple. Al tender

$$\rho \rightarrow 0, \quad V(\bar{x}_s) = \frac{\sigma^2}{n} (1 + (n-1)*0) = \frac{\sigma^2}{n} = V(\bar{x}).$$

Podemos entonces definir tres tipos de poblaciones con base a lo antes señalado:

Población aleatoria

Población ordenada

Población periódica

**Definición:** Una población es aleatoria si sus elementos están ordenados al azar.

Los elementos de una muestra sistemática seleccionados de una población aleatoria se espera que sean heterogéneos con un  $\rho$  aproximadamente igual a cero. Por lo tanto

cuando  $N$  es grande la varianza de  $\bar{x}_s$  es aproximadamente igual a la varianza de  $\bar{x}$  basada en el muestreo aleatorio simple. El muestreo sistemático en este caso es equivalente al muestreo aleatorio simple. Es este tipo de poblaciones las que estaremos interesados en estudiar ya que las estimaciones se realizan con el mismo cuerpo de fórmulas que para el muestreo aleatorio simple.

**Definición:** Una población es ordenada si los elementos dentro de la población están ordenados en magnitud de acuerdo con algún esquema.

Una muestra sistemática extraída de una población ordenada es generalmente

heterogénea con  $\rho \leq 0$ . Cuando  $N$  es grande  $\rho \leq 0$  y,  $V(\bar{x}_s) \leq V(\bar{x})$ . Donde  $V(\bar{x}_s)$

es la varianza por muestreo sistemático y  $V(\bar{x})$  es la varianza por muestreo aleatorio simple.

Por lo tanto una muestra sistemática de una población ordenada proporciona más información que una muestra aleatoria simple por unidad de costo, debido a que la

varianza de  $\bar{x}_s$  es menor que la varianza correspondiente de  $\bar{x}$ .

**Definición:** Una población es periódica si los elementos de la población tienen variación cíclica.

Los elementos de una muestra sistemática extraída de una población periódica pueden ser homogéneas (esto es  $\rho > 0$ ). Cuando  $N$  es grande y  $\rho > 0$   $V(\bar{x}_s) > V(\bar{x})$ . Por lo

tanto, en este caso el muestreo sistemático proporciona menos información que el muestreo aleatorio por unidad de costo.

#### 4. Estimaciones basadas en una muestra aleatoria sistemática para poblaciones aleatorias

En base a lo estudiado en el subtema anterior, las estimaciones de la media poblacional, total poblacional y proporción poblacional del muestreo aleatorio sistemático para poblaciones aleatorias, son las mismas formulas que se aplican al muestreo aleatorio simple para las estimaciones de la media, el total poblacional y la proporción poblacional. Es importante insistir que esto es exclusivamente para poblaciones aleatorias, no consideramos ni poblaciones ordenadas ni poblaciones periódicas para las estimaciones de interés, debido a que las varianzas son mayores o menores respecto al muestreo aleatorio simple. De encontrarnos poblaciones de ese tipo es conveniente proceder a desordenarlas conforme lo sugerido en el subtema 2.

#### 5. Estimador de la media poblacional $\mu$ , varianza estimada de la media y límite para el error de estimación de la media

Retomando las formulas del muestreo aleatorio simple tenemos:

Estimador de la media poblacional  $\mu$ , varianza estimada de la media y límite para el error de estimación de la media serán:

El estimador de la media poblacional  $\mu$ ,  $\mu = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

varianza estimada de  $\bar{x}$ :  $\hat{v}(\bar{x}) = \frac{s^2}{n} \left( \frac{N-n}{N} \right)$

donde:  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}$

Límite para el error de estimación:  $z\sqrt{\hat{v}(\bar{x})} = z\sqrt{\frac{s^2}{n} \left( \frac{N-n}{N} \right)}$

donde  $\frac{N-n}{N}$  es un factor de corrección por población finita.

**Veamos el siguiente ejercicio:**

La sección de control de calidad de una empresa usa el muestreo sistemático para estimar la cantidad promedio de llenado en latas de 12 onzas que sale de una línea de producción. Los datos de la tabla adjunta representan una muestra sistemática de 1-en-50 de la producción de un día. Estime  $\mu$ , y establezca un límite para el error de estimación. Suponga que  $N = 1800$ .

Cantidad de llenado (en onzas)					
12.00	11.97	12.01	12.03	12.01	11.80
11.91	11.98	12.03	11.98	12.00	11.83
11.87	12.01	11.98	11.87	11.90	11.88
12.05	11.87	11.91	11.93	11.94	11.89
11.72	11.93	11.95	11.97	11.93	12.05
11.85	11.98	11.87	12.05	12.02	12.04

**Solución:**

Debemos estimar  $\mu$  y establecer el límite para el error de estimación

Los datos que se nos suministra son aparte de los valores de llenado en onzas:

$$N = 1,800 \quad n = 36$$

$$\bar{x} = \frac{\sum_{i=1}^{36} x_i}{n} = \frac{430.01}{36} = 11.9447222$$

Entonces:

Varianza estimada de  $\bar{x}$ :

$$\hat{v}(\bar{x}) = \frac{s^2}{n} \left( \frac{N-n}{N} \right)$$

como  $s^2$  no es conocida, se calcula de los datos de la tabla que se nos ha suministrado, resultando:

$$s^2 = 0.00616849 \text{ verifíquelo!}$$

$$\text{Donde } \hat{v}(\bar{x}) = \frac{0.00616849}{36} \left( \frac{1800-36}{1800} \right) = 0.0001679$$

El límite para el error de estimación con un 95% de confianza, está dado por:

$$Z\sqrt{\hat{v}(\bar{x})} = 1.96\sqrt{0.0001679} = 0.0254$$

Por lo que el intervalo donde encontramos a  $\mu$  con un nivel de confianza del 95% está dado por:

$$\mu \rightarrow \bar{x} \pm Z\sqrt{\hat{v}(\bar{x})} = 11.9447 \pm 0.0254$$

$$\mu \rightarrow (11.9193, 11.9701)$$

La cantidad promedio de llenado de las latas que salen de la línea de producción, se estima entre 11.9193 y 11.9701 onzas con un nivel de confianza del 95%.

## 6. Estimador del total poblacional $\tau$ , varianza estimada del total y límite para el error de estimación del total poblacional serán:

El estimador del total poblacional  $\tau$ ,  $\hat{\tau} = N\bar{x} = \frac{N \sum_{i=1}^n x_i}{n}$

La Varianza estimada de  $\tau$ :  $\hat{v}(\hat{\tau}) = \hat{v}(N\bar{x}) = N^2 \left( \frac{s^2}{n} \right) \left( \frac{N-n}{N} \right)$

Donde  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$

El Límite para el error de estimación:  $z\sqrt{\hat{v}(N\bar{x})} = z\sqrt{N^2 \left( \frac{s^2}{n} \right) \left( \frac{N-n}{N} \right)}$

### Veamos el siguiente ejercicio:

Un colegio está interesado en mejorar sus relaciones con su comunidad vecina. Una muestra sistemática de 1-en-150 de los  $N = 4500$  estudiantes listados en el directorio es tomada para estimar el total de dinero gastado en ropa durante un trimestre del año escolar. Los resultados de la muestra están listados en la tabla anexa. Use los datos para estimar  $\tau$ , y establezca un límite para el error de estimación.

Estudiante	Cantidad gastada (en córdobas)	Estudiante	Cantidad gastada (en córdobas)
1	30	16	32
2	22	17	14
3	10	18	29
4	62	19	48
5	28	20	50
6	31	21	9
7	40	22	15
8	29	23	6
9	17	24	93
10	51	25	21
11	29	26	20
12	21	27	13
13	13	28	12
14	15	29	29
15	23	30	38

**Solución:** Se requiere estimar  $\tau$  y establecer un límite para el error de estimación. El estimador del total poblacional  $\tau$ , está dado por:

$$\hat{\tau} = N\bar{x}$$

donde  $\bar{x} = 28.333 \dots$ , verifiquel o de la tabla de cantidad gastada, como  $N = 4,500$

$$\hat{\tau} = (4,500)(28.333 \dots) = 127,500$$

La varianza estimada de  $\tau$ , está dada por:

$$\hat{v}(\hat{\tau}) = N^2 \left( \frac{s^2}{n} \right) \left( \frac{N-n}{N} \right) \text{ donde } s^2 \text{ debe calcularse de la tabla que se nos da.}$$

$$s^2 = 338.643678 \text{ verifiquel o!}$$

$$\hat{v}(\hat{\tau}) = (4500)^2 \left( \frac{338.643678}{30} \right) \left( \frac{4,500 - 30}{4,500} \right) = 227,060.586.1$$



El límite para el error de estimación está dado por:

$$Z\sqrt{v(\tau)} = 1.96\sqrt{227,060.586.1} = 29,534.32 \quad \text{para un nivel de confianza del 95\%}$$

El intervalo de confianza de  $\tau$  estará dado por :

$$\tau \rightarrow \hat{\tau} \pm Z\sqrt{v(\tau)} = 127,500 \pm 29,534.32$$

$$\tau \rightarrow (97,965.68 ; 157,034.32)$$

El total de dinero gastado en ropa durante un trimestre del año escolar de la comunidad vecina, se estima entre 97,965.68 y 157,034.32 unidades monetarias con un nivel de confianza del 95%.

**7. El estimador de la proporción poblacional  $p$ , varianza estimada de la proporción y límite para el error de estimación de la proporción poblacional serán:**

Estimador de la proporción poblacional  $p$ :

$$\hat{p} = \frac{x}{n} : \frac{\text{Número de éxitos en la muestra}}{\text{Número muestreado}}$$

Varianza estimada de  $\hat{p}$  :  $v(\hat{p}) = \frac{\hat{p}\hat{q}}{n-1} \left( \frac{N-n}{N} \right)$  donde:  $\hat{q} = 1 - \hat{p}$

Límite para el error de estimación:  $z\sqrt{v(\hat{p})} = z\sqrt{\frac{\hat{p}\hat{q}}{n-1} \left( \frac{N-n}{N} \right)}$

**Veamos el siguiente ejercicio:**

La gerencia de una compañía privada está interesada en estimar la proporción de empleados que favorecen una nueva política de inversión. Una muestra sistemática de 1-en-10 es obtenida de los empleados que salen del edificio al final de un día de trabajo en particular. Use los datos de la tabla adjunta para estimar  $p$ , la proporción a favor de la nueva política y establezca un límite para el error de estimación. Suponga  $N = 2000$ .

Empleado Muestreado	Respuesta
3	1
13	0
23	1
.	.
.	.
.	.
1993	1
$\sum_{i=1}^{200} x_i = 132$	

**Solución:**

$$p = ?$$

$$k = 10$$

$$N = 2,000$$

como se ha realizado una muestra sistemática de  $n = 200$ , en dicha muestra, la proporción de empleados a favor, está dada por :

$$\bar{p} = \frac{\sum_{i=1}^{200} x_i}{n} = \frac{132}{200} = 0.66$$

La varianza estimada de  $\bar{p}$  es :

$$v(\bar{p}) = \frac{\bar{p} \bar{q}}{n-1} \left( \frac{N-n}{N} \right)$$

$$v(\bar{p}) = \frac{(0.66)(0.34)}{200-1} \left( \frac{2000-200}{2000} \right) = 0.001$$

Por lo que el límite para el error de estimación para un Nivel de confianza del 95% está dado por :

$$Z\sqrt{v(\bar{p})} = 1.96\sqrt{0.001} = 0.062$$

Entonces  $p$  se encuentra en el intervalo (0.598, 0.722)

La proporción se estima entre 59.8% y 72.2% de los empleados de la compañía que favorecen la nueva política de inversión con un nivel de confianza del 95%.

## 8. Cálculo del tamaño de la muestra para la media poblacional, el total poblacional y la proporción poblacional

### a. Tamaño de la muestra para estimar la media poblacional ( $\mu$ )

Tamaño de muestra requerido para estimar  $\mu$  con un límite para el error de estimación B:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \quad \text{donde } D = \frac{B^2}{Z^2}$$

### **Veamos el siguiente ejercicio:**

Use los datos del ejercicio que resolvimos relacionado a la sección de control de calidad de una empresa que está interesada en la cantidad promedio de llenado en latas de 12 onzas, para determinar el tamaño de muestra requerido para estimar  $\mu$  dentro de 0.03 unidades.

En este caso  $B = 0.03$  y se requiere encontrar  $n$ . 
$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \quad \text{donde } D = \frac{B^2}{Z^2}$$

El valor de  $D$  lo encontramos, para un nivel de confianza del 95%,  $Z = 1.96$ . 
$$D = \frac{(0.03)^2}{(1.96)^2} = 0.00023428$$

como  $\sigma^2$  no es conocida, se asume  $s^2$  que nos dará un valor de  $n$  aproximado, sustituyendo los datos conocidos se tiene:

$$n = \frac{1800(0.00616849)}{(1800-1)(0.00023428) + 0.00616849} = \frac{11.103282}{0.42763821} = 25.96$$

Por lo que  $n = 26$

*El tamaño de la muestra que se requiere para realizar la estimación de la media poblacional del llenado de latas de 12 onzas por parte de la sección de control de calidad es de  $n = 26$  con un error de estimación de 0.03 y con un nivel de confianza del 95%.*

## b. Tamaño de la muestra para estimar el total poblacional ( $\tau$ )

Tamaño de muestra requerido para estimar  $\tau$  con un límite para el error de estimación B:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \quad \text{donde } D = \frac{B^2}{Z^2 N^2}$$

**Veamos el siguiente ejercicio:**

*¿Qué tamaño de muestra es necesario para estimar  $\tau$  en el anterior ejercicio del colegio interesado en el gasto de ropa por trimestre, con un límite para el error de estimación aproximadamente igual a C\$ 10,000? ¿Qué esquema de muestreo sistemático recomendaría?*

*Se requiere encontrar  $n$ , para estimar  $\tau$  con un límite para el error de estimación de  $B = 10,000$ .*

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \quad \text{donde } D = \frac{B^2}{Z^2 N^2}$$

$$\text{calculando } D = \frac{(10,000)^2}{(1.96)^2 (4,500)^2} = 1.285472617$$

*como  $\sigma^2$  no se conoce, se asume el valor calculado de  $s^2$  en el ejercicio anterior, obteniendo:*

$$n = \frac{4,500(338.643678)}{(4,500-1)(1.285472617) + 338.643678} = \frac{1,523,896.55}{6,121.984982} = 248.92 \quad \text{donde } n = 249$$

*El tamaño de muestra que se requiere para estimar el total de dinero gastado en ropa durante un trimestre del año escolar para la comunidad vecina, es de  $n = 249$ , con un error de estimación de C\$ 10,000 y con un nivel de confianza del 95%.*

*El esquema de muestreo sistemático que se recomienda es de 1-en-18, para cumplir la condición*

$$k \leq \frac{N}{n} = \frac{4500}{249} = 18.07, \text{ por lo que } k = 18.$$

## c. Tamaño de la muestra para estimar la proporción poblacional ( $p$ )

Tamaño de muestra requerido para estimar  $p$  con un límite para el error de estimación B:

$$n = \frac{Npq}{(N-1)D + pq} \quad \text{donde } q = 1 - p \text{ y } D = \frac{B^2}{Z^2}$$

**Veamos el siguiente ejercicio:**

Para la situación referida en el anterior ejercicio de la proporción de empleados que favorecen una nueva política de inversión, determine el tamaño de muestra requerido para estimar  $p$ , con un límite para el error de estimación de 0.01 unidades. ¿Qué tipo de muestra sistemática deberá obtenerse?

Si asumimos que el límite para el error de estimación es 0.01, esto es  $B = 0.01$

El cálculo del tamaño de la muestra se realizará entonces por la fórmula:

$$n = \frac{Npq}{(N-1)D + pq} \text{ donde } D = \frac{B^2}{Z^2}$$

Asumiendo un 95% de Nivel de confianza obtenemos:

$$D = \frac{(0.01)^2}{(1.96)^2} = 0.00002603$$

$$n = \frac{2000(0.66)(0.34)}{1999(0.00002603) + (0.66)(0.34)} = \frac{448.8}{0.05203561 + 0.2244} = \frac{448.88}{0.27643561} = 1,623.52$$

$$n = 1,624$$

El tamaño de la muestra que se requiere para estimar la proporción poblacional de empleados que favorecen una nueva política de inversión, con un error de estimación de 0.01 unidades y con un nivel de confianza del 95% es de  $n = 1,624$ . El tipo de muestra sistemática que deberá de obtenerse es de 1-en-1.

### Actividad de autoaprendizaje No. 3

1. En una comunidad se realiza un censo. Además de la información usual que se obtiene de la población, los investigadores preguntan a los habitantes de cada vigésima casa cuánto tiempo la han habitado. Estos resultados se resumen a continuación:

$$n = 115$$

$$N = 2300$$

$$\sum x_i = 407.1(\text{años})$$

$$\sum x_i^2 = 2011.15$$

Uso estos datos para estimar la cantidad promedio de tiempo que las personas han vivido en su casa actual. Establezco un límite para el error de estimación.

2. Los funcionarios de un museo están interesados en el número total de personas que visitan el lugar durante un periodo de 180 días cuando una costosa colección de antigüedades está en exhibición. Puesto que el control de visitantes en el museo cada día es muy costoso, los funcionarios deciden obtener estos datos cada décimo día. La información de esta muestra sistemática de 1 en 10 se resume en la tabla adjunta.

Día	Número de personas que visitan el museo
3	160
13	350
23	225
.	.
.	.
.	.
173	290
$\sum_{i=1}^{18} x_i = 4,868$	
$\sum_{i=1}^{18} x_i^2 = 1,321,450$	

Uso estos datos para estimar  $\tau$ , el número total de personas que visitan el museo durante el periodo específico. Establezco un límite para el error de estimación.

Encontraré las respuestas a esta actividad de autoaprendizaje en la página 183 al final de la unidad autoformativa II.

## **Resumen final de la unidad autoformativa II**

En la segunda unidad autoformativa hemos estudiado el diseño de muestras y dos de los cuatro métodos de muestreo probabilístico, el aleatorio simple y el aleatorio sistemático.

En el tema "A", hemos profundizado en el concepto de diseño de muestras, su importancia y significado. Estudiamos el error de estimación en el muestreo y las estimaciones puntual y por intervalo así como el tamaño de la muestra. En este primer tema tratamos de establecer conexión con los contenidos estudiados en el módulo de Estadística, hemos establecido la importancia de las estimaciones y con ello las posibilidades de realizar la inferencia estadística de los hallazgos en muestras representativas y probabilísticas a la población, este aspecto es la esencia de una de las ramas más importante de la estadística, denominada "estadística inferencial".

De los otros dos temas (B y C), hemos estudiado los dos métodos de muestreo probabilístico, el aleatorio simple y el aleatorio sistemático.

Definimos cada método, estableciendo el procedimiento de selección de la muestra y procediendo a establecer las estimaciones de la media, el total y la proporción poblacional, así como el cálculo de las varianzas y los límites del error de estimación y cálculo del tamaño de las muestras.

Planteamos la aplicabilidad del muestreo aleatorio simple en investigaciones sobre poblaciones pequeñas, plenamente identificables, constituidas en universos específicos y diferenciados, debido a la dificultad de poseer de forma completa el marco muestral apropiado para poblaciones grandes.

En el caso del muestreo aleatorio sistemático, estudiamos que es una variante del método aleatorio simple que difiere en la forma de selección de las unidades muestrales. Posee como ventajas el extender la muestra sobre toda la población, es de fácil aplicación pero se deberá tener cuidado en el ordenamiento de la población dado que si nos encontramos con poblaciones ordenadas o periódicas, debemos desordenar sus elementos para así proceder a su aplicación y con ello, su varianza sea equivalente a la del muestreo aleatorio simple.

Con lo aprendido en la presente unidad autoformativa tenemos los elementos complementarios para desarrollar nuestro trabajo final, según lo requerido por las orientaciones para su elaboración. En él aplicaremos cualquiera de los dos métodos estudiados. Los otros dos los analizaremos en la tercera unidad autoformativa.

## Evaluación Final de la unidad autoformativa II

La base de datos que se anexa, refleja las notas finales de una población de estudiantes que han cursado la asignatura x en un cuatrimestre en la Universidad Centroamericana.

Se le solicita:

1. Seleccionar una muestra aleatoria simple de tamaño  $n = 20$ , utilizando la tabla de números aleatorios que se anexa en su unidad autoformativa.
2. Realizar el cálculo de la media muestral (nota media). Considerar que se ha aplicado una prueba piloto y se ha encontrado que  $s^2 = 300$
3. Estimar la varianza de la media y el límite para el error de estimación de la media con un nivel de confianza del 95%.
4. Establecer el intervalo de confianza para la media poblacional.
5. Comparar el valor real obtenido de la media poblacional con el intervalo de confianza encontrado en el punto anterior. ¿Qué puede concluir?
6. Calcular el tamaño de la muestra para la media poblacional con un nivel de confianza del 95% y con un error máximo permisible de 5 puntos.
7. Qué conclusiones y enseñanzas le ha permitido esta evaluación?

Base de datos población Nota Final de la asignatura X

Caso	Sexo	Nota Final
1	F	46
2	M	63
3	F	49
4	F	89
5	F	96
6	F	26
7	M	78
8	M	85
9	F	60
10	F	87
11	F	80
12	F	83
13	M	62
14	F	84
15	M	62
16	F	92
17	F	81
18	F	71
19	F	77
20	F	33
21	F	85
22	M	59
23	F	70
24	F	65
25	F	51
26	F	91
27	M	33
28	F	95
29	F	64
30	F	98
31	F	20
32	F	58
33	F	73
34	F	66
35	M	91
36	F	90
37	M	75
38	F	91
39	F	73
40	M	86
41	F	82
42	M	72
43	M	79
44	F	91
45	F	83
46	F	81
47	F	84
48	F	83
49	F	71
50	F	73
51	F	73
52	F	92
53	M	75
54	F	73
55	F	71
56	F	62
57	F	78

Caso	Sexo	Nota Final
58	M	81
59	F	84
60	F	98
61	F	96
62	F	83
63	F	76
64	F	96
65	F	81
66	M	89
67	F	86
68	F	82
69	F	89
70	F	53
71	F	54
72	F	92
73	M	30
74	M	98
75	F	79
76	F	72
77	M	77
78	F	78
79	F	80
80	F	76
81	F	94
82	F	94
83	F	84
84	F	77
85	F	81
86	F	26
87	M	75
88	F	73
89	M	73
90	F	90
91	F	74
92	F	88
93	F	70
94	F	73
95	F	71
96	F	92
97	F	71
98	F	77
99	F	75
100	F	62
101	F	70
102	F	71
103	F	70
104	M	73
105	F	76
106	F	24
107	M	13
108	M	92
109	M	86
110	M	65
111	M	80
112	F	67
113	M	90
114	M	72

Caso	Sexo	Nota Final
115	F	42
116	F	58
117	M	60
118	M	80
119	M	38
120	M	13
121	F	65
122	F	70
123	F	71
124	F	65
125	F	95
126	M	60
127	F	96
128	M	74
129	F	94
130	F	61
131	F	61
132	M	63
133	M	60
134	F	85
135	F	47
136	F	95
137	M	77
138	F	71
139	F	95
140	M	48
141	F	90
142	M	65
143	F	85
144	F	85
145	F	86
146	M	60
147	F	60
148	M	62
149	M	88

Media Poblacional:	<b>73.05369</b>
Varianza Poblacional:	<b>318.01054</b>



## Hoja de respuestas

### Respuesta a la Actividad de Autoaprendizaje No. 1

1. El objetivo que se persigue es la de realizar una inferencia sobre el parámetro poblacional a partir de los estimadores muestrales y el máximo error que se está dispuesto a tolerar al realizar dicha estimación.
2. El error de estimación se define como la diferencia absoluta que se presenta, al comparar el estimador muestral y el parámetro poblacional. Si el estimador muestral es 11.5 y el parámetro poblacional 12, entonces el valor del error de estimación es:  $|11.5 - 12| = 0.5$ , entonces el valor de  $B = 0.5$ .
3. Los factores que se deben elegir para el cálculo del tamaño de la muestra, son tres:
  - El grado de confianza, el cual se asocia al valor de  $Z$  que determinará en alguna medida el ancho del intervalo de confianza que deseamos permitir.
  - El máximo error permisible ( $B$ ), el cual es el tamaño máximo del error que estaremos dispuestos a aceptar al calcular el tamaño de la muestra.
  - La variación de la población, esta se asocia a la variabilidad de la población objeto de estudio. Como esta no es conocida normalmente, conviene realizar una prueba piloto que nos permita encontrar la varianza de la prueba piloto ( $s^2$ ), y utilizarla o sustituirla por la varianza poblacional de la característica de interés que estemos estudiando. Recordar que para el caso del tamaño muestral en el caso de la media:
 
$$n = \left( \frac{Z * S}{B} \right)^2$$
4.  $\mu \rightarrow 55 \pm (2.58) * \frac{10}{\sqrt{49}}; \quad \mu \rightarrow (51.32 ; 58.69)$
5.  $\bar{p} = 0.625 \quad \rho = \bar{p} \pm Z \sqrt{\frac{p^*(1-p)}{n}} \quad \rho \rightarrow (0.5776 ; 0.6724)$
6. a.  $\bar{p} = 0.625$   
 b.  $\rho = \bar{p} \pm Z \sqrt{\frac{p^*(1-p)}{n}} \quad \rho \rightarrow (0.5776 ; 0.6724)$   
 c.  $B = Z * \sigma_{\bar{p}} \quad \therefore n = 196$

### Respuestas a la Actividad de Autoaprendizaje No. 2

1. El número de muestras es también de 10 muestras

$$\sigma^2 = 2$$

$$V(\bar{x}) = E(\bar{x} - \mu)^2 = \sum (\bar{x} - \mu)^2 p(\bar{x}) = \frac{1}{3}$$

$$V(\bar{x}) = \frac{5-3}{5-1} \times \left( \frac{2}{3} \right) = \frac{1}{3}$$

Con lo que se demuestra lo solicitado.

$$z\sqrt{V(\bar{x})} = 1.96\sqrt{12.3948} = 6.9$$

2.  $\mu \rightarrow (\bar{x} - z\sqrt{V(\bar{x})}, \bar{x} + z\sqrt{V(\bar{x})})$   
 $\mu \rightarrow (5.6, 19.4)$

3. a. 420  
 b. 490

- c. Ambos son insesgados  
d. Estimación realizada en el inciso b).

$$\begin{aligned} \bar{p} &= \frac{120}{200} = 0.60 \\ 4. \quad N \bar{p} &= 2000 \times 0.60 = 1,200 \\ z \sqrt{V(\bar{p})} &= 129 \\ \tau &\rightarrow (1071, 1329) \end{aligned}$$

$$\begin{aligned} \bar{p} &= \frac{11}{60} = 0.183 \\ 5. \quad z \sqrt{V(\bar{p})} &= 0.107 \\ p &\rightarrow (0.076, 0.290) \end{aligned}$$

$$6. \quad n = 86$$

$$\begin{aligned} \bar{x} &= \frac{40.1}{8} = 5.013 \\ 7. \quad s^2 &= 1.647 \\ z \sqrt{V(\bar{x})} &= 0.852 \\ \mu &\rightarrow (4.161, 5.865) \end{aligned}$$

8. El diseño de muestreo es de manera individual, teniendo como elementos a ser seleccionados los siguientes:
- Marco muestral que se desprende de la población objetivo
  - Diseño de muestreo que se elegirá el procedimiento aleatorio simple y que se le aplicará a la población correspondiente.
  - El grado de confianza del estimador
  - El máximo error permisible
  - La variabilidad de la población que se calculará con la aplicación de la prueba piloto.
  - El tamaño de la muestra que se calculará con los tres elementos anteriores y con base a la fórmula del tamaño de la muestra para el muestreo aleatorio simple.
  - La selección de los elementos que pasan a formar parte de la muestra y que se eligen por medio del procedimiento aleatorio simple.

### Respuestas a la Actividad de Autoaprendizaje No. 3

$$\begin{aligned} \bar{x} &= \frac{407.1}{115} = 3.54 \text{ años} \\ 1. \quad z \sqrt{V(\bar{x})} &= 0.4 \\ \mu &\rightarrow (3.14, 3.94) \\ 2. \quad \bar{\tau} &= 180 \times 270.44 = 48680 \\ z \sqrt{V(N\bar{x})} &= 1,348.82 \end{aligned}$$

## Glosario de términos

**Coeficiente de elevación:** Número que resulta de dividir el tamaño de la población respecto el tamaño de la muestra en una muestra sistemática y que al ser sumado al número aleatorio originalmente elegido nos permite elegir o generar el resto de elementos de la muestra. Debe de ser un entero que cumpla con la condición:  $K \leq \frac{N}{n}$ .

**Diseño de muestras:** Se le denomina diseño de muestras, al método o procedimiento que se sigue para elegir de los miembros o elementos de una población los que serán parte de la muestra.

**Distribución de probabilidad:** Es la enumeración de todos los resultados de un experimento estadístico junto con la probabilidad asociada a cada uno.

**Elemento:** Se refiere a un objeto o individuo de una población o muestra del cual se toma una medición.

**Error de muestreo:** Diferencia entre un estadístico muestral y el parámetro poblacional correspondiente.

**Estadístico muestral:** Es una característica medible de una muestra.

**Estimador:** Es una función de variables aleatorias observables y quizás otras constantes conocidas, usado para estimar un parámetro poblacional.

**Estimador por intervalo:** Es el intervalo de valores que se estima en el que probablemente se encuentra un parámetro poblacional, con base en información muestral.

**Estimador puntual:** Es un estadístico muestral puntual que se utiliza para estimar un parámetro poblacional.

**Grado de confianza:** Valor de  $z$  en una distribución normal estandarizada que resulta al encontrar el área bajo la curva normal  $(1 - \alpha)$  y permite calcular el intervalo de confianza ( Para un intervalo de confianza del 95% , esto es  $(1 - 0.05)$ ,  $z = 1.96$ ).

**Inferencia Estadística:** Proceso de deducir conclusiones generales acerca de un grupo completo (población) con base en información estadística obtenida de un grupo pequeño (la muestra).

**Intervalo de confianza:** Es el intervalo dentro del que se espera probabilísticamente, esté un parámetro poblacional al realizar una estimación por intervalo

**Insesgado:** Se denomina insesgado a un estadístico muestral, cuando el valor esperado del estadístico muestral es igual al parámetro poblacional que se estima.

**Límite para el error de estimación:** Es el tamaño del posible error de estimación que se calcula multiplicando el grado de confianza por la raíz cuadrada de la varianza estimada de una estimación ( ejemplo para la media  $\bar{x} : \pm z \sqrt{V(\bar{x})}$  ).

**Máximo error permisible:** Es el máximo error tolerable por parte del investigador para un nivel de confianza específico, al aplicar las técnicas del muestreo.

**Monoetápico:** Procedimiento de extracción de la muestra en un solo paso.

**Muestreo aleatorio simple:** Acción de extraer muestras mediante el uso del azar, de tal forma que todos los elementos de la población tengan la misma probabilidad de formar parte de la muestra.

**Muestreo aleatorio sistemático:** Acción de extraer muestras mediante un criterio racional, seleccionándose al azar un punto de inicio y después se elige cada k-ésimo elemento de la población para la muestra.

**Parámetro poblacional:** Es una característica medible de una población.

**Polietápico:** Procedimiento de extracción de la muestra en dos o más pasos.

**Unidad de muestreo:** Colecciones disjuntas de elementos de la población que permiten formar un marco muestral.

**Varianza estimada:** Nivel de variabilidad de un estimador, que nos permite calcular el límite para el error de estimación.

**Sesgo:** Una consecuencia posible si a ciertos miembros de la población se les niega la oportunidad de ser seleccionados en la muestra. Como resultado, la muestra puede no ser representativa de la población.

## **Bibliografía**

Mason D. B. y Lind D. A. Estadística para administración y economía 2000. Editorial Alfaomega, octava edición, México.

Berenson M. L. y Levine D.M. Estadística básica en Administración, Conceptos y aplicaciones. 1992. Editorial PHH, cuarta edición, México.

Sheaffer R., Mendenhal W., y Ott L. Elementos de muestreo. 1987. Editorial Iberoamericana, tercera edición, México.

Koroliuk V. S. Manual de la teoría de probabilidades y estadística matemática. 1981. Editorial Mir, primera edición, URSS

Zajarov V. K., Sevastianov B. A. Y Chistiakov V. P. Teoría de las probabilidades. 1983. Editorial Mir, primera edición, URSS.

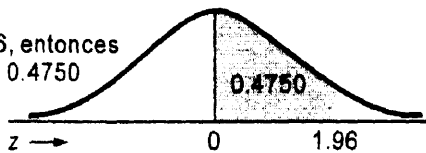
Cohran W. G. Técnicas de muestreo. 1995. Editorial CECSA, primera edición, México.

Mendenhall W. Y Reinmuth J. Estadística para administración y economía. 1995. Editorial Iberoamericana, tercera edición, México.

## Apéndice Unidad Autoformativa II

### Anexo 1

### Tabla de Distribución Normal Estandarizada

Áreas bajo la curva normal										
<p>Ejemplo: si <math>z = 1.96</math>, entonces <math>P(0 \leq z) = 0.4750</math></p> 										
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Tomada de: Mason y Lind, Estadística para Administración y Economía. 1998. Página 968

**Anexo 2**  
**Tabla de Números Aleatorios del 1 al 99999**

92513	21156	20618	78167	87337	80703	00468	84107	58120	15456
17743	07198	90623	93143	58624	99172	90774	32892	43015	03868
64861	63732	82018	99240	11925	87715	08917	64625	51627	20766
79728	09753	43999	23550	20290	45512	06476	50545	49641	68162
20515	97350	04927	87226	11553	40866	70782	20944	65761	08406
55469	10628	09666	09906	94742	11203	99330	38092	26386	38890
95834	67434	10367	25356	03517	11773	12159	61605	79374	73538
38344	94791	86959	90375	39792	31626	16217	75227	59118	48758
58805	62440	80292	69705	80900	53725	19260	63853	52996	41111
13310	73409	14683	72271	83141	31214	06020	36270	89706	59971
63713	76433	82594	13916	49596	51822	19609	27454	85470	89592
74104	73061	81653	70740	30044	11351	97125	92733	73051	84786
78861	42464	71587	29797	69687	98987	73885	91051	56021	55387
36915	77574	29370	77071	38475	39808	26999	61025	74503	48624
04771	12920	35438	84635	57077	71178	05781	95589	19444	50655
38792	98613	08548	65459	46643	31589	26608	93722	71904	98448
22358	87444	47701	58366	28319	66264	90515	16041	62258	35767
85925	79736	54218	88132	33224	19743	29210	26962	79052	44391
52972	54548	85857	60644	12414	63024	30719	47093	57483	98950
01504	34410	87397	45152	13270	70784	40097	74700	36362	27201
47310	81172	43366	95102	12440	66368	71484	84735	78198	65864
52773	71303	43874	02350	50081	87145	08226	45971	87660	77714
10822	41517	94528	27204	34618	21898	92037	77282	00834	91593
45122	47428	78536	16203	47494	40969	75182	44961	04057	75998
57860	98766	70665	22204	50186	92901	43092	01484	94901	38907
97394	86975	63503	83789	66147	18745	98056	56883	09887	68377
04824	69980	08281	08365	81162	14938	15126	14297	97930	87086
72933	93269	88155	37800	46676	45646	93489	83874	82111	62150
16305	67104	66060	28206	53165	79230	01372	53440	27867	09628
02674	21739	07376	51503	39818	77886	66043	53012	66435	65773
72311	39822	42810	19800	96401	27412	36409	69962	93899	11187
20706	35672	86437	61921	65572	92677	95636	57183	58219	63017
42718	18888	63674	90418	40960	08617	24701	31349	72595	34379
24010	42118	44494	68694	33997	26586	86540	52297	74910	56444
32067	87027	19144	18467	40376	35642	05817	27684	19042	03323
44569	68369	59771	39055	17045	64723	65400	66571	75542	63935
08487	63190	89557	44757	52447	62902	87699	50010	63877	96156
85834	44712	33304	54909	75046	49907	64365	42403	61689	72287
05150	00751	15552	87821	72474	73283	88592	61628	12753	78396
17670	88611	90535	04493	23554	07028	19741	00796	12694	65857
05221	65192	51547	52682	81566	94180	78802	47901	76064	54259
20977	74751	03998	78282	64292	48037	09946	02345	61880	45268
57906	12627	93767	06761	55512	53175	26354	73948	55615	53165
28352	33164	93269	93833	58574	62083	13382	00109	46278	49877
94165	27864	64602	68693	53619	64668	07040	33622	33066	40964
56693	42385	00601	65723	54104	63014	05768	07673	88769	42281
09131	74506	63078	30882	59738	70331	63669	16305	60160	72148
17652	95540	87088	89126	16730	79871	48653	37999	25753	89591
63576	61256	93052	49509	39740	39844	73818	30614	29605	62099
40912	46784	97819	67113	23390	79137	50734	33970	71863	79725

Tomado de: Hayes Bob E.. ¿Cómo medir la satisfacción del cliente. 1999. Página 252

**Unidad Autoformativa III**  
**"Muestreo probabilístico aleatorio**  
**estratificado y aleatorio por**  
**conglomerado"**



## **Presentación**

*A partir de esta unidad autoformativa, estudiaremos los dos últimos métodos de muestreo probabilísticos, de los cuatro que tiene previsto, esta asignatura. Estos métodos son el aleatorio estratificado y el aleatorio por conglomerado.*

*Iniciaremos cada tema con la definición del método de muestreo, posteriormente estudiaremos el procedimiento de selección de la muestra, para concluir con la presentación de diversas fórmulas que nos ayudarán a realizar las estimaciones de la media poblacional, total poblacional, proporción poblacional y el tamaño de la muestra.*

*En esta unidad requerimos que usted, identifique y aplique correctamente las fórmulas para realizar los cálculos apropiados.*

## **Objetivos de la unidad autoformativa III**

1. Aplico adecuadamente los procedimientos de selección de muestras aleatorias estratificadas y por conglomerados.
2. Aplico las formulas adecuadas para las estimaciones de la media poblacional, total poblacional, proporción poblacional y tamaño de muestras aleatorias estratificadas y por conglomerados

## Contenido de la unidad autoformativa III

### Tema A. Muestreo aleatorio estratificado

1. Definición
2. Procedimiento de selección de una muestra aleatoria estratificada
3. Estimaciones basadas en una muestra aleatoria estratificada
4. Estimador de la media poblacional, varianza estimada de la media y limite para el error de estimación de la media
5. Estimador del total poblacional, varianza estimada del total y limite para el error de estimación del total
6. Estimador de la proporción población, varianza estimada de la proporción y limite para el error de estimación de la proporción poblacional
7. Calculo del tamaño de la muestra para las medias, total poblacional y proporción poblacional.

### Tema B. Muestreo aleatorio por conglomerado

1. Definición
2. Procedimiento de selección de una muestra aleatoria por conglomerado
3. Estimaciones basadas en una muestra aleatoria por conglomerado
4. Estimador de la media poblacional, varianza estimada de la media y limite para el error de estimación de la media
5. Estimador del total poblacional, varianza estimada del total y limite para el error de estimación del total
6. Cálculo del tamaño de la muestra para las medias y total poblacional
7. Estimador de una proporción población, varianza estimada de la proporción y limite para el error de estimación de la proporción poblacional
8. Calculo del tamaño de la muestra para las proporciones de la poblacional.

## A. MUESTREO ALEATORIO ESTRATIFICADO

### 1. Definición

Una muestra aleatoria estratificada, es un muestra aleatoria que se obtiene separando los elementos de la población en grupos disjuntos, llamados estratos, y seleccionando una muestra aleatoria simple dentro de cada estrato.

La partición o fraccionamiento de la población marco del estudio en subdivisiones constituye la base del muestreo estratificado. En él cada unidad del universo pertenece a una sola subdivisión, que llamamos estrato, y el conjunto de los estratos, subdivisiones que se realizan del universo, constituyen la base sobre la que opera el proceso de muestreo.

#### A manera de ejemplo:

Si nuestra población es la del grupo de estudiantes que hemos tomado como ejemplo en la unidad autoformativa anterior,  $N = 50$ . Supongamos entonces que es nuestro interés estratificar dicha población de acuerdo al sexo, obtendremos dos estratos con tamaños por ejemplo  $N_1=30$  damas y  $N_2=20$  caballeros. Observar que los miembros que pertenecen al estrato "damas", no pertenecen al estrato "caballeros" o viceversa, esto es que los grupos en que se subdivide la población son grupos disjuntos que le denominaremos estratos.

#### a. Ventajas del muestreo aleatorio estratificado

Las principales ventajas de este tipo de muestreo, ampliamente utilizado, son las siguientes:

- Permite tratar de forma independiente a cada uno de los estratos. Esto facilita la utilización de diferentes métodos de muestreo así como la estimación, por separado, de ciertas subpoblaciones constituidas en dominios de estudio.
- Se pueden reducir las varianzas de las estimaciones muestrales.
- Aumenta la precisión de las estimaciones.
- Facilita la coordinación de los trabajos de campo.
- El costo de la recolección y el análisis de los datos se reduce al estratificar en grupos cuyos elementos tienen características similares pero que difieren de un grupo a otro.
- Usualmente la varianza del estimador de la media poblacional se reduce usando muestreo aleatorio estratificado, debido a que la variabilidad dentro de los estratos es generalmente menor que la variabilidad de la población.
- Se obtienen estimadores separados para los parámetros de cada estrato, sin necesidad de seleccionar otra muestra e incurrir en mayores gastos.

- La estratificación en los procesos de muestreo ofrece una gran flexibilidad ya que posibilita la utilización simultánea, en una misma muestra, de distintos métodos de muestreo en función de las necesidades del diseño y de la información disponible sobre cada estrato. Facilita, además, la estimación por separado de los distintos estratos, siempre que la muestra este adecuadamente dimensionada. Esto abre el camino a la desagregación del universo en dominios diferenciados, posibilitando el análisis de cada uno de ellos.
- Por otra parte, la estratificación, al reunir en cada estrato a unidades homogéneas entre sí y heterogéneas en relación con las de los otros estratos, contribuye a reducir la varianza de las estimaciones, lo que se traduce en ganancias en precisión, es decir, en disminución de los errores de muestreo. Esto quiere decir que, si se pudiera hacer la estratificación ideal, este tipo de muestreo sería más preciso que el aleatorio simple. En la práctica, esto significa que, para un error muestral dado, es necesario hacer menor número de entrevistas si se utiliza el muestreo estratificado que si se utiliza el aleatorio simple. Por otra parte, si se conocen las varianzas por estrato se pueden aplicar fracciones de muestreo distintas en cada estrato de acuerdo con el valor de la varianza. En estratos más homogéneos se podrían utilizar fracciones de muestreo más pequeñas que en estratos más heterogéneos.
- Finalmente, la estratificación puede facilitar la coordinación de los trabajos de campo, ya que permite separar la muestra por estratos. En este caso, cada estrato o cada varios estratos pueden formar áreas operativas distintas bajo la responsabilidad del Jefe de zona que coordina y controla la recogida de la información.

## **b. Criterios de estratificación**

Haciendo referencia ahora a los criterios de estratificación, hay que tener en cuenta las variables que se van a utilizar y el número de estratos resultante.

- En cuanto a lo primero: "las variables utilizadas para la estratificación deberían estar correlacionadas con las variables objeto de la investigación". Es más, si fuera posible, debería recurrirse a variables de alta correlación que garantizaran la homogeneidad de los estratos. Así, para hacer un diseño muestral para el estudio de la pobreza, se podría partir de variables de ingresos, igual que, para hacer un estudio sobre el nivel de conocimientos de la población, habría que utilizar variables referidas a los estudios realizados.
- En lo que respecta al número de estratos, se puede señalar que "no es aconsejable llevar la división muy lejos", ya que "los estratos muy pequeños contribuyen muy poco a las ganancias de la estratificación" y "la formación de sólo unos cuantos estratos producirá típicamente la mayoría de las ganancias posibles a partir de una variable".

Aunque la teoría, en este campo, es conocida, sin embargo, resulta difícil estratificar por variables altamente correlacionadas, puesto que no se suelen conocer las varianzas. En general, y como criterio más viable, se suele recurrir a variables espaciales: comunidades autónomas, departamentos, municipios, etcétera, o a subdivisiones inherentes al universo

en estudio. Así, el universo alumnos de universidad, se puede estratificar por universidades, facultades, especialidades, cursos, etcétera; a los profesores de Educación Media y Primaria, por tipo de enseñanza pública o privada, comunidades autónomas, tamaño del centro, etcétera. En estos casos también es posible controlar el número de estratos, siendo el investigador el que tiene que decidir las variables que va a utilizar para estratificar y el número de estratos resultantes. En cualquiera de los casos, y como norma general, hay que dejar claro que la estratificación tiene un límite y que, incluso operativamente, la multiplicación excesiva de estratos puede complicar el diseño, ya que aparecerán estratos vacíos o casi vacíos.

Este tipo de muestreo es el más utilizado tanto en muestreos monoetápicos como en polietápicos, en los que suele acompañar a otras formas de muestreo. De ahí que se debiera dedicar, en cada caso, una especial atención a la estratificación, clave de las ventajas de este sistema de muestreo.

## **2. Procedimiento de selección de una muestra aleatoria estratificada**

**Primero:** Elegir claramente la población objetivo de estudio, definiendo con ello el criterio de clasificación de conformación de los estratos.

Es indispensable tener claramente definida la población objetivo a la cual le estimaremos los parámetros, para ello el primer paso del procedimiento de selección de la muestra estratificada se asocia tanto a tener definida la población objetivo como también a definir el criterio de estratificación.

Hemos anteriormente ejemplificado con la población de estudiantes de tamaño  $N = 50$  y de acuerdo al criterio "sexo" hemos obtenido  $N_1 = 30$  damas y  $N_2 = 20$  caballeros.

El criterio de estratificación debe asociarse a lo que estemos estudiando, es así que podemos la misma población estratificarla de acuerdo al departamento de procedencia, de acuerdo al tipo de trabajo que realizan los estudiantes, etc.

Es importante en este primer paso establecer claramente la población objetivo y el criterio de estratificación para permitir adecuadamente los siguientes pasos.

**Segundo:** Especificar clara y detalladamente cada estrato, asociando a cada elemento de la población con uno y solo un estrato. En este segundo paso es que se procede a ubicar a cada miembro de la población objetivo en el estrato correspondiente, así todas las damas de la población pasarán a formar parte del estrato "Damas" y todos los caballeros pasaran al estrato "Caballeros", llegando a conformarse los dos estratos con tamaño  $N_1 = 30$  para el estrato Damas y  $N_2 = 20$  para el estrato Caballeros.

En el caso de que a esa misma población estemos interesados estratificarla en Departamento de procedencia ¿Qué haríamos? Favor prepararse para compartir su opinión con el grupo.

**Tercero:** Una vez especificados los estratos, se procede a seleccionar una muestra aleatoria simple en cada estrato, en base al procedimiento estudiado en el muestreo aleatorio simple. Ya definidos los estratos y ordenados los elementos que lo forman, se tendrá entonces que aplicar el muestreo aleatorio simple a cada estrato. Así que en el

caso del estrato  $N_1=30$ , se tendrá que sacar la muestra aleatoria de tamaño  $n_1$ ; para el estrato  $N_2=20$ , se tendrá que sacar la muestra de tamaño  $n_2$ . Las muestras al ser sacadas se procederá como si cada estrato es una población siguiendo el procedimiento del muestreo aleatorio simple ya estudiado.

En esto debemos estar seguros de que las muestras seleccionadas en los estratos sean independientes. Esto es, se deben aplicar diferentes esquemas de muestreo aleatorio dentro de cada estrato, de tal manera que las observaciones elegidas en un estrato no dependan de las que se han elegido en otro estrato.

Pero aún nos queda la pregunta, **¿De qué tamaño serán las muestras de los estratos?, es decir, ¿de qué tamaño será  $n_1$  y  $n_2$ ?**

Si seguimos asumiendo el tamaño de muestra  $n$  para la población entonces  $n = n_1 + n_2$ , ¿qué valor tendrá  $n_1$  y  $n_2$ ? Observemos que podemos elegir varios tamaños de  $n_1$  y  $n_2$ , podría ser que  $n_1 = n_2 = 5$ , que  $n_1=2$  y  $n_2=8$ , que  $n_1=6$  y  $n_2=4$ , etc.

A la relación de los tamaños de muestra de los estratos, versus los tamaños de estrato y de la población se le conoce con el nombre de afijación.

Notemos que para el caso de  $n_1=6$  y  $n_2=4$ , se cumple que  $N_1/n_1 = N_2/n_2 = 5$  para estos tamaños de muestra la afijación es proporcional al tamaño de los estratos. En el caso de que  $n_1 = n_2 = 5$ , este tipo de afijación se le denomina con el nombre de afijación simple. Estudiaremos otras relaciones de tamaño en base a otros criterios más adelante en la presente unidad autoformativa.

### ***Veamos un ejemplo de como se opera en el muestreo estratificado:***

*Universo: Trabajadores de Iberia.*

*Estratificación: Dada la gran heterogeneidad de funciones que existen dentro de la compañía se pensó en subdividir al personal en estratos homogéneos. Para ello, y a partir del catálogo de puestos de trabajo de la empresa, un grupo de personas conocedoras de la misma hizo los siguientes estratos:*

- *Directivos y técnicos de grado superior.*
- *Técnicos de grado medio y técnicos auxiliares.*
- *Administrativos.*
- *Especialistas y servicios auxiliares.*
- *Pilotos y operadores de radio.*
- *Auxiliares de vuelo.*

*Se hicieron, por lo tanto, seis estratos, cada uno de los cuales reunía a personas de intereses profesionales parecidos y distintos de los intereses de los otros grupos. Aparte de las ventajas de esta estratificación para mejorar la precisión de las estimaciones, también permitió la realización de seis submuestras. Dichas submuestras son representativas de cada una de las poblaciones de las que fueron extraídas.*

*Tamaño de la muestra: El tamaño de la muestra se fijo en 1.500 entrevistas, buscando un compromiso entre representatividad del universo y de cada uno de los grupos. La muestra es plenamente representativa del universo trabajadores de Iberia y, con menor precisión, de cada uno de los grupos en que se dividió el universo.*

*Afijación: La afijación no fue proporcional porque se buscaban submuestras representativas. De ahí que se fijara el número mínimo de entrevistas por estrato en 200 y el máximo en 300, dada la limitación de 1.500 entrevistas en total.*

*Ponderación: Cada una de las submuestras se puede tabular por separado. Si se desea tabular la muestra conjuntamente es necesario proceder a la ponderación porque las tasas de muestreo son distintas en cada submuestra.*

*Selección de los elementos de la muestra: La selección se hizo en una sola etapa utilizando la técnica del muestreo sistemático. Para ello se partió de los listados nominales de la población de cada estrato, dentro del cual se hizo la selección de la submuestra correspondiente. Primero se eligió, al azar, un número no superior al coeficiente de elevación y, posteriormente, se eligieron los números siguientes, con sólo añadir al anterior el valor del coeficiente de elevación. Una alternativa a esta solución hubiera sido extraer los elementos de cada submuestra de la lista nominal de individuos, utilizando el muestreo aleatorio simple.*

### 3. Estimaciones basadas en una muestra aleatoria estratificada

Inicialmente estableceremos la notación que será utilizada en el muestreo aleatorio estratificado:

- $L$  : Número de estratos.  
 $\bar{X}_{est}$  : Media en el muestreo estratificado.  
 $\bar{X}_i$  : Media del estrato  $i$ .  
 $N$  : Tamaño de la población, número de unidades muestrales en la población.  
 $N_i$  : Tamaño del estrato  $i$ , número de unidades muestrales en el estrato  $i$ .  
 $n_i$  : Tamaño de la muestra del estrato  $i$ .  
 $s_i^2$  : varianza de la muestra del estrato  $i$ .  
 $\hat{V}(\bar{X}_{est})$  : varianza estimada de la media estratificada.

Observemos que:  $N = N_1 + N_2 + \dots + N_L$

Estimador de la media poblacional  $\mu$ , varianza estimada de la media y límite para el error de estimación de la media

El estimador de la media poblacional  $\mu$  lo calcularemos por medio de:

$$\bar{X}_{est} = \frac{1}{N} [N_1 \bar{X}_1 + N_2 \bar{X}_2 + \dots + N_L \bar{X}_L] = \frac{1}{N} \sum_{i=1}^L N_i \bar{X}_i, \text{ fórmula 3.1}$$

La Varianza estimada de  $\bar{X}_{est}$  se encuentra con la fórmula siguiente:

$$\hat{V}(\bar{X}_{est}) = \frac{1}{N^2} \sum_{i=1}^L N_i^2 \left( \frac{N_i - n_i}{N_i} \right) \left( \frac{S_i^2}{n_i} \right), \text{ fórmula 3.2}$$

El Límite para el error de estimación con:

$$Z \sqrt{\hat{V}(\bar{X}_{est})} = Z \sqrt{\frac{1}{N^2} \sum_{i=1}^L N_i^2 \left( \frac{N_i - n_i}{N_i} \right) \left( \frac{S_i^2}{n_i} \right)}, \text{ fórmula 3.3}$$

**Veamos el siguiente ejercicio:**

Una escuela desea estimar la calificación promedio que puede ser obtenida en un examen de comprensión de lectura por estudiantes de sexto grado. Los estudiantes de la escuela son agrupados en tres estratos, los que aprenden rápido en el estrato I y los que aprenden lento en el estrato III. La escuela decide esta estratificación porque de esta manera se reduce la variabilidad en las calificaciones del examen. El sexto grado contiene 55 estudiantes en el estrato I, 80 en el estrato II y 65 en el estrato III. Una muestra aleatoria estratificada de 50 estudiantes es asignada proporcionalmente y produce muestras irrestrictas aleatorias de  $n_1 = 14$ ,  $n_2 = 20$  y  $n_3 = 16$  de los estratos I, II y III. El examen se aplica a la muestra de estudiantes y se obtienen los resultados que se muestran en la tabla. Estime la calificación promedio para este grado y establezca un límite para el error de estimación.

Estrato I		Estrato II		Estrato III	
80	92	85	82	42	32
68	85	48	75	36	31
72	87	53	73	65	29
85	91	65	78	43	19
90	81	49	69	53	14
62	79	72	81	61	31
61	83	53	59	42	30
		68	52	39	32
		71	61		
		59	42		

**Solución:**

Para calcular  $\bar{x}_{est}$  aplicamos la fórmula 3.1, realizamos primero el cálculo de las  $\bar{x}_i$  y calculamos además las  $s_i^2$ :

$$\begin{array}{llll}
 \bar{x}_1 = 79.72 & S_1^2 = 105.1429 & n_1 = 14 & N_1 = 55 \\
 \bar{x}_2 = 64.75 & S_2^2 = 158.1974 & n_2 = 20 & N_2 = 80 \\
 \bar{x}_3 = 37.44 & S_3^2 = 186.1292 & n_3 = 16 & N_3 = 65 \quad N = 200
 \end{array}$$

El estimador de la media poblacional  $\mu$  resulta:

$$\begin{aligned}
 \bar{x}_{est} &= \frac{1}{N} \sum_{i=1}^L N_i \bar{x}_i \\
 \bar{x}_{est} &= \frac{1}{200} (55 \times 79.72) + (80 \times 64.75) + (65 \times 37.44) = \frac{1}{200} (4,384.6 + 5,180 + 2,433.6) \\
 \bar{x}_{est} &= \frac{1}{200} (11,998.2) = 59.991
 \end{aligned}$$

La Varianza estimada de  $\bar{x}_{est}$  la calculamos con la fórmula 3.2:

$$\begin{aligned}
 \hat{v}(\bar{x}_{est}) &= \frac{1}{N^2} \sum_{i=1}^L N_i^2 \left( \frac{N_i - n_i}{N_i} \right) \left( \frac{S_i^2}{n_i} \right) \\
 \hat{v}(\bar{x}_{est}) &= \frac{1}{(200)^2} \left[ (55)^2 \left( \frac{55 - 14}{55} \right) \left( \frac{105.1429}{14} \right) + (80)^2 \left( \frac{80 - 20}{80} \right) \left( \frac{158.1974}{20} \right) + (65)^2 \left( \frac{65 - 16}{65} \right) \left( \frac{186.1292}{16} \right) \right] \\
 &= \frac{1}{(200)^2} [16935.52 + 37967.376 + 37,051.343] = \frac{1}{(200)^2} (91954.24) = 2.299
 \end{aligned}$$

$$\hat{v}(\bar{x}_{est}) = 2.299$$

$$\sqrt{\hat{v}(\bar{x}_{est})} = 1.516197875$$

El límite para el error de estimación para  $Z = 1.96$  con nivel de confianza del 95% se calcula aplicando la fórmula 3.3

$$Z \sqrt{\hat{v}(\bar{x}_{est})} = 1.96 \times 1.516197875 = 2.97$$

$$\mu \rightarrow \bar{x}_{est} \pm Z \sqrt{\hat{v}(\bar{x}_{est})} = 59.991 \pm 2.97$$



Por lo tanto la estimación de la media poblacional  $\mu$  de la calificación en el examen de comprensión de lectura, con el nivel del 95% de confianza, se encuentra entre 57 y 63 puntos.

#### 4. Estimador del total poblacional $\tau$ , varianza estimada del total y límite para el error de estimación del total poblacional

El estimador del total poblacional  $\tau$  se calcula por:  $\hat{\tau} = N\bar{X}_{est} = \sum_{i=1}^L Ni\bar{X}_i$ , fórmula 3.4

Varianza estimada para  $\hat{\tau} = N\bar{X}_{est}$  por medio de:  $\hat{V}(N\bar{X}_{est}) = \sum_{i=1}^L Ni^2 \left( \frac{Ni - ni}{Ni} \right) \left( \frac{Si^2}{ni} \right)$ , fórmula 3.5

El Límite para el error de estimación:  $Z\sqrt{\hat{V}(N\bar{X}_{est})} = Z\sqrt{\sum_{i=1}^L Ni^2 \left( \frac{Ni - ni}{Ni} \right) \left( \frac{Si^2}{ni} \right)}$ , fórmula 3.6

#### Veamos el siguiente ejercicio:

Una corporación desea estimar el número total de horas-hombre perdidas debido a accidentes de los empleados en un mes determinado. Ya que los obreros, los técnicos y los administrativos tienen diferentes tasas de accidentes, el investigador decide usar muestreo aleatorio estratificado, con cada grupo formando un estrato. Estime el número total de horas-hombre perdidas durante el mes indicado y establezca un límite para el error de estimación. Use los datos de la tabla acompañante, obtenida de una muestra piloto de 18 obreros, 10 técnicos y 2 administrativos.

I (Obreros)		II (Técnicos)		III (Administrativos)	
8	0	24	4	5	1
0	32	16	0	24	8
6	16	0	8	12	
7	4	4	3	2	
9	8	5	1	8	
18	0	2			

Los tamaños de los estratos son:  $N_1 = 132$ ,  $N_2 = 92$  y  $N_3 = 27$

#### Solución:

Debemos estimar el total poblacional  $\tau$  aplicando la fórmula 3.4:

Los tamaños de las muestras son:  $n_1=18$ ,  $n_2=10$  y  $n_3=2$

$$\begin{aligned} N\bar{x}_{est} &= \sum_{i=1}^L Ni\bar{x}_i \\ \bar{x}_1 &= \frac{8+0+6+7+9+18+0+32+16+4+8+0+24+16+0+4+5+2}{18} = \frac{159}{18} = 8.8 \\ \bar{x}_2 &= \frac{4+0+8+3+1+5+24+12+2+8}{10} = \frac{67}{10} = 6.7 \\ \bar{x}_3 &= \frac{8+1}{2} = 4.5 \\ N\bar{x}_{est} &= (132 \times 8.8) + (92 \times 6.7) + (27 \times 4.5) = 1,899.5 \end{aligned}$$

La Varianza estimada de  $N\bar{x}_{est}$  se calcula aplicando la fórmula 3.5:

$$v(N\bar{x}_{est}) = \sum_{i=1}^L N_i^2 \left( \frac{N_i - n_i}{N_i} \right) \left( \frac{S_i^2}{n_i} \right)$$

Encontrando las varianzas de los estratos por medio de los datos suministrados en la t5tabla, se tiene:

$$S_1^2 = \frac{2,791 - 1,404.5}{17} = 81.5588$$

$$S_2^2 = \frac{903 - 448.9}{9} = 50.4556$$

$$S_3^2 = \frac{65 - 40.5}{1} = 24.5 \quad \text{¡ Favor verificarlo!}$$

La Varianza estimada para el total  $\tau$  es entonces:

$$\begin{aligned} v(N\bar{x}_{est}) &= (132)^2 \left( \frac{132 - 18}{132} \right) \left( \frac{81.5588}{18} \right) + (92)^2 \left( \frac{92 - 10}{92} \right) \left( \frac{50.4556}{10} \right) + (27)^2 \left( \frac{27 - 2}{27} \right) \left( \frac{24.5}{2} \right) = \\ &= 68,183.1568 + 38,063.7046 + 8,268.75 = 114,515.6114 \end{aligned}$$

El límite para el error de estimación se calcula con la fórmula 3.6:

$$Z\sqrt{v(N\bar{x}_{est})} = Z\sqrt{\sum_{i=1}^L N_i^2 \left( \frac{N_i - n_i}{N_i} \right) \left( \frac{S_i^2}{n_i} \right)}$$

Para el 95% de nivel de confianza,  $Z = 1.96$

$$Z\sqrt{v(N\bar{x}_{est})} = 1.96(338.4016) = 663.267$$

$$\tau \rightarrow N\bar{x}_{est} \pm z\sqrt{V(\bar{x}_{est})} = 1,899.5 \pm 663.26$$

Por lo que la estimación del total de horas perdidas  $\tau$  debido a accidentes de los empleados se encuentra con un nivel de confianza del 95% entre 1,236.2 y 2,562.8 horas-hombres en el mes indicado.

## 5. Estimador de la proporción poblacional $p$ , varianza estimada de la proporción y límite para el error de estimación de la proporción poblacional

El estimador de la proporción poblacional  $p$  se calcula por:  $\hat{p}_{est} = \frac{1}{N} \sum_{i=1}^L N_i \hat{p}_i$ , fórmula 3.7

La Varianza estimada de  $\hat{p}_{est}$  por medio de:  $\hat{V}(\hat{p}_{est}) = \frac{1}{N^2} \sum_{i=1}^L N_i^2 \left( \frac{N_i - n_i}{N_i} \right) \left( \frac{\hat{p}_i \hat{q}_i}{n_i - 1} \right)$ ,

fórmula 3.8

El Límite para el error de la estimación es entonces:  $Z\sqrt{\hat{V}(\hat{p}_{est})} = Z\sqrt{\frac{1}{N^2} \sum_{i=1}^L N_i^2 \left( \frac{N_i - n_i}{N_i} \right) \left( \frac{\hat{p}_i \hat{q}_i}{n_i - 1} \right)}$ ,

fórmula 3.9

**Veamos el siguiente ejercicio:**

Una cadena de almacenes de departamentos está interesada en estimar la proporción de cuentas por cobrar negligentes. La cadena consiste de 4 almacenes. Así que el costo de muestreo es reducido. Se usa muestreo aleatorio estratificado, con cada tienda como un estrato. Ya que no se dispone de información referente a las proporciones poblacionales antes del muestreo, se usa la asignación proporcional. De la tabla acompañante, estime  $p$ , la proporción de cuentas negligentes para la cadena, y fije un límite para el error de estimación.

	Estrato I	Estrato II	Estrato III	Estrato IV
Número de cuentas por cobrar	$N_1=65$	$N_2=42$	$N_3=93$	$N_4=25$
Tamaño de muestra	$n_1=14$	$n_2=9$	$n_3=21$	$n_4=6$
Número muestral de cuentas negligentes	4	2	8	1

**Solución:**

Debemos estimar la proporción poblacional  $P$ :

El número total de cuentas por cobrar es  $N = 225$

La proporción estratificada se estima por medio de la fórmula 3.7:

$$\hat{P}_{est} = \frac{1}{N} \sum_{i=1}^4 N_i \hat{P}_i$$

$$\text{Estrato I : } \hat{P}_1 = \frac{4}{14} = 0.2857$$

$$\text{Estrato II : } \hat{P}_2 = \frac{2}{9} = 0.2222$$

$$\text{Estrato III : } \hat{P}_3 = \frac{8}{21} = 0.3809$$

$$\text{Estrato IV : } \hat{P}_4 = \frac{1}{6} = 0.1667$$

$$\begin{aligned} \hat{P}_{est} &= \frac{1}{225} [(65 \times 0.2857) + (42 \times 0.2222) + (93 \times 0.3809) + (25 \times 0.1667)] \\ &= \frac{1}{225} (18.57 + 9.33 + 35.42 + 4.16) = \frac{1}{225} (67.48) = 0.2999 \end{aligned}$$

La Varianza estimada de  $\hat{P}_{est}$  se calcula por medio de la fórmula 3.8:

$$\begin{aligned} v(\hat{P}_{est}) &= \frac{1}{N^2} \left[ N_1^2 v(\hat{P}_1) + N_2^2 v(\hat{P}_2) + N_3^2 v(\hat{P}_3) + N_4^2 v(\hat{P}_4) \right] = \frac{1}{N^2} \sum_{i=1}^4 N_i^2 \left( \frac{N_i - n_i}{N_i} \right) \left( \frac{\hat{p}_i \hat{q}_i}{n_i - 1} \right) \\ v(\hat{P}_{est}) &= \frac{1}{(225)^2} \left[ (65)^2 \left( \frac{65 - 14}{65} \right) \left( \frac{0.2857 \times 0.7143}{13} \right) + (42)^2 \left( \frac{42 - 9}{42} \right) \left( \frac{0.2222 \times 0.7778}{8} \right) + \right. \\ &\quad \left. (93)^2 \left( \frac{93 - 21}{93} \right) \left( \frac{0.3809 \times 0.6191}{20} \right) + (25)^2 \left( \frac{25 - 6}{25} \right) \left( \frac{0.1667 \times 0.8333}{5} \right) \right] = \\ &= \frac{1}{225^2} [52.0392 + 29.9423 + 78.9509 + 13.1965] = \frac{1}{225^2} (174.1289) \\ &= 0.003439 \end{aligned}$$

$$\sqrt{v(\hat{P}_{est})} = 0.0586$$

El Límite para el error de estimación se calcula por la fórmula 3.9:

$$Z \sqrt{v(\hat{P}_{est})} = 1.96 (0.0586) = 0.1149$$

$$p \rightarrow \bar{p}_{est} \pm z \sqrt{v(\bar{p}_{est})} = 0.2999 \pm 0.1149$$

La proporción de cuentas por cobrar negligentes  $p$  para la cadena de almacenes de departamentos, se estima entre el 18.5% y el 41.48% con un nivel del 95% de confianza.

## 6. Cálculo del tamaño de la muestra para la media poblacional, el total poblacional y la proporción poblacional

El tamaño de muestra aproximada que se requiere para estimar  $\mu$  o  $\tau$  con un límite  $B$  para el error de estimación, se calcula por:

$$n = \frac{\sum_{i=1}^L N_i^2 \sigma_i^2 / w_i}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}, \text{ fórmula 3.10}$$

donde  $w_i$  se conoce y es la fracción de observaciones asignados al estrato  $i$ ,  $\sigma_i^2$  es la varianza poblacional para el estrato  $i$ .  $w_i = \frac{n_i}{n}$ ,  $i=1,2,\dots,L$ ; y

$$D = \frac{B^2}{Z^2} \text{ para estimar } \mu \qquad D = \frac{B^2}{Z^2 N^2} \text{ para estimar } \tau$$

Recordemos que en el subtema B, habíamos comentado el tamaño de las muestras respecto al tamaño de los estratos. Al elegir el tamaño de muestra  $n$ , existen muchas maneras para dividir  $n$  entre los tamaños de muestra de los estratos individuales  $n_1, n_2, n_3, \dots, n_L$ . Cada división puede originar una varianza diferente para la media muestral. También hemos planteado que en el muestreo se persigue el objetivo de obtener una cantidad especificada de información a un costo mínimo, es por tal razón que la asignación que establezcamos estará influida por tres factores que son:

- 1) *El número total de elementos en cada estrato.* En este sentido, se tendrá que asignar tamaños grandes de muestras a los estratos que contienen gran número de elementos y asignar tamaños pequeños de muestra a los estratos con pocos elementos.
- 2) *La variabilidad de las observaciones dentro de cada estrato.* Al poseer estratos heterogéneos, requeriremos de una asignación de muestra de tamaño grande, en cambio si poseemos estratos homogéneos, asignaremos muestras de tamaño pequeño.
- 3) *El costo por obtener una observación de cada estrato.* Si el costo para obtener una observación varía de un estrato a otro, se tomarán muestras pequeñas a los estratos con altos costos y muestras grandes a los estratos con costos bajos. Lo anterior lo basamos en la necesidad de que se mantengan los costos totales de la muestra al mínimo posible, brindándonos el mayor nivel de información para nuestro estudio.

En ese sentido, la asignación aproximada que minimiza el costo para el valor fijo de  $v(\bar{X}_{est})$  o que minimiza  $v(\bar{X}_{est})$  para un costo fijo se determina por medio de:

$$n_i = n \left( \frac{N_i \sigma_i / \sqrt{c_i}}{\sum_{k=1}^L N_k \sigma_k / \sqrt{c_k}} \right), \text{ fórmula 3.11}$$

donde  $N_i$  denota el tamaño del  $i$ -ésimo estrato,  $\sigma_i^2$  la varianza poblacional para el  $i$ -ésimo estrato y  $c_i$  el costo para obtener una observación individual del  $i$ -ésimo estrato.

El tamaño de muestra aproximada anteriormente planteado en la fórmula 3.10 se nos transforma en:

$$n = \frac{\left( \sum_{k=1}^L N_k \sigma_k / \sqrt{c_k} \right) \left( \sum_{i=1}^L N_i \sigma_i \sqrt{c_i} \right)}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}, \text{ fórmula 3.12}$$

**Veamos el siguiente ejercicio:**

Una corporación desea obtener información acerca de la efectividad de una máquina comercial. Se va a entrevistar por teléfono a un número de jefes de división, para pedirles que califiquen la maquinaria en base en una escala numérica. Las divisiones están localizadas en Norteamérica, Europa y Asia. Es por eso que se usa muestreo estratificado. Los costos son mayores para las entrevistas de los jefes de división localizados fuera de Norteamérica. La tabla siguiente proporciona los costos por entrevista, varianzas aproximadas de las calificaciones, y los  $N_i$ , que se han establecido.

La corporación quiere estimar la calificación promedio con  $V(\bar{x}_{est}) = 0.1$ . Elija el tamaño de muestra  $n$  que obtiene este límite y encuentre la asignación apropiada.

Estrato I (Norteamérica)	Estrato II (Europa)	Estrato III (Asia)
$C1=\$9$	$C2=\$25$	$C3=\$36$
$\sigma_1^2=2.25$	$\sigma_2^2=3.24$	$\sigma_3^2=3.24$
$N_1=112$	$N_2=68$	$N_3=39$

**Solución:** Como  $Z \sqrt{v(\bar{x}_{est})} = B$

Para  $Z = 1.96$  corresponde el 95% de confianza y calculando el valor de  $D$ , se tiene:

$$\begin{aligned} v(\bar{x}_{est}) &= 0.1, \text{ entonces} \\ B &= 1.96 \sqrt{0.1} = 0.6198 \\ D &= \frac{B^2}{Z^2} \text{ para el promedio} \\ D &= \frac{(0.6198)^2}{(1.96)^2} = 0.1 \end{aligned}$$

El tamaño de muestra se calcula por medio de la fórmula 3.12  
Se tiene que :  $N = N_1 + N_2 + N_3 = 112 + 68 + 39 = 219$

$$n = \frac{\left( \sum_{k=1}^L \frac{N_k \sigma_k}{\sqrt{C_k}} \right) \left( \sum_{i=1}^L N_i \sigma_i \sqrt{C_i} \right)}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}$$

$$\sum_{k=1}^L \frac{N_k \sigma_k}{\sqrt{C_k}} = (112) \frac{1.5}{3} + (68) \frac{1.8}{5} + (39) \frac{1.8}{6} = 56 + 24.48 + 11.7 = 92.18$$

$$\sum_{i=1}^L N_i \sigma_i \sqrt{C_i} = (112)(1.5)(3) + (68)(1.8)(5) + (39)(1.8)(6) = 1,537.2$$

$$\sum_{i=1}^L N_i \sigma_i^2 = (112)(2.25) + (68)(3.24) + (39)(3.24) = 598.68$$

$$N^2 D = (219)^2 (0.1) = 4,796.1$$

$$n = \frac{(92.18)(1,537.2)}{4,796.1 + 598.68} = \frac{141,699.0960}{5394.78} = 26.266$$

$$n = 26$$

La asignación apropiada se calcula por medio de la fórmula 3.11 y será:

$$n_i = n \left( \frac{\frac{N_i \sigma_i}{\sqrt{C_i}}}{\sum_{k=1}^L \frac{N_k \sigma_k}{\sqrt{C_k}}} \right)$$

Como conocemos que  $n = 26$  y  $\sum_{k=1}^L \frac{N_k \sigma_k}{\sqrt{C_k}}$  lo hemos calculado anteriormente y nos resultó un valor de 92.18, calculamos los valores de  $n_i$ :

$$n_1 = \frac{26}{92.18} \frac{(112)(1.5)}{\sqrt{9}} = 15.79 \cong 16$$

$$n_2 = \frac{26}{92.18} \frac{(68)(1.8)}{\sqrt{25}} = 6.90 \cong 7$$

$$n_3 = \frac{26}{92.18} \frac{(39)(1.8)}{\sqrt{36}} = 3.3 \cong 3$$

La asignación apropiada para estimar la calificación promedio con  $V(\bar{x}_{est}) = 0.1$ , resulta ser:  $n_1 = 16, n_2 = 7, n_3 = 3; n = 26$  relacionada a la calificación de la efectividad de la máquina comercial de interés de la corporación. Lo anterior con un nivel de confianza del 95%.

Observar que las fórmulas 3.11 y 3.12 se han aplicado cuando los factores:

Número total de elementos por cada estrato	¡Son diferentes para todos los estratos!
Varianza en los estratos y	
Costos unitarios por estratos	

Si el costo para obtener información es el mismo para todos los estratos, esto es  $c_1 = c_2 = \dots = c_L$ , entonces las fórmulas 3.11 y 3.12 se nos transforman respectivamente en las fórmulas 3.13 y 3.14:

$$n_i = n \left( \frac{N_i \sigma_i}{\sum_{i=1}^L N_i \sigma_i} \right), \text{ fórmula 3.13}$$

$$n = \frac{\left( \sum_{i=1}^L N_i \sigma_i \right)^2}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}, \text{ fórmula 3.14}$$

Este método de seleccionar  $n_1, n_2, \dots, n_L$  se denomina asignación de Neyman.

Observemos que en la asignación de Neyman, los factores que varían entre los estratos son:

- 1) Número total de elementos para cada estrato y
- 2) Varianza en los estratos.

El factor costo unitario por estrato es el mismo para la obtención de una observación en cualquier estrato.

### **Veamos el siguiente ejercicio:**

Una corporación desea estimar el número total de horas-hombre perdidas debido a accidentes de los empleados, en un mes determinado. Ya que los obreros, los técnicos y los administrativos tienen diferentes tasas de accidentes, el investigador decide usar muestreo aleatorio estratificado, con cada grupo formando un estrato. Datos de años previos sugieren las varianzas mostradas en la tabla anexa, para el número de horas-hombre perdidas por empleado en los tres grupos, y de datos actuales se obtienen los tamaños de los estratos. Determine la asignación de Neyman para una muestra de  $n = 30$  empleados.

<b>I (Obreros)</b>	<b>II (Técnicos)</b>	<b>III (Administrativo)</b>
$\sigma_1^2=36$	$\sigma_2^2=25$	$\sigma_3^2=9$
$N_1=132$	$N_2=92$	$N_3=27$

### **Solución:**

Recordemos que este ejercicio se nos planteó en un momento anterior, ahora se nos presenta información de las varianzas de los estratos que utilizaremos para encontrar los valores de tamaño de muestras por medio de la asignación de Neyman, aplicando la fórmula 3.13:

$$n_i = n \left( \frac{N_i \sigma_i}{\sum_{i=1}^L N_i \sigma_i} \right)$$

como  $n = 30$ , se debe determinar  $n_i$ :

$$n_1 = 30 \left( \frac{132 \times 6}{132 \times 6 + 92 \times 5 + 27 \times 3} \right) = 30 \times \frac{792}{1333} = 17.8 \approx 18$$

$$n_2 = 30 \left( \frac{92 \times 5}{132 \times 6 + 92 \times 5 + 27 \times 3} \right) = 30 \times \frac{460}{1333} = 10.3 \approx 10$$

$$n_3 = 30 \left( \frac{27 \times 3}{132 \times 6 + 92 \times 5 + 27 \times 3} \right) = 30 \times \frac{81}{1333} = 1.8 \approx 2$$

La asignación de Neyman corresponde a:  $n_1 = 18, n_2 = 10, n_3 = 2$  para las muestras de los estratos, de una muestra total de  $n = 30$  con el propósito de estimar el número total de horas-hombre perdidas debido a accidentes de los empleados.

---

Recordar que cuando el factor costo unitario es el mismo para todos los estratos, mientras la varianza y el número de elementos por estrato es diferente, se deben aplicar las fórmulas 3.13 y 3.14. La asignación que resulta de la aplicación de la fórmula 3.13, es denominada con el nombre de asignación de Neyman.

Además de los costos iguales, si se tiene que las varianzas también son iguales, las ecuaciones anteriores 3.13 y 3.14 se simplifican, esto es: la 3.13 por la 3.15 y la 3.14 por la 3.16 de tal manera que:

$$n_i = n \left( \frac{N_i}{N} \right), \text{ fórmula 3.15}$$

Este método de asignación de la muestra a los estratos es llamado asignación proporcional dado que los tamaños de las muestras  $n_1, n_2, \dots, n_L$  son proporcionales a los tamaños de los estratos  $N_1, N_2, \dots, N_L$ .

En asignación proporcional, la expresión para el cálculo de  $n$  (fórmula 3.14), se convierte en:

$$n = \frac{\sum_{i=1}^L N_i \sigma_i^2}{ND + \frac{1}{N} \sum_{i=1}^L N_i \sigma_i^2}, \text{ fórmula 3.16}$$

En el caso de las proporciones, el tamaño de muestra aproximado que se requiere para estimar  $p$ , con un límite  $B$  para el error de estimación, estarán dados por:

$$n = \frac{\sum_{i=1}^L N_i^2 p_i q_i / w_i}{N^2 D + \sum_{i=1}^L N p_i q_i}, \text{ fórmula 3.17}$$

donde  $w_i$  es la fracción de observaciones asignadas al estrato  $i$ ,  $p_i$  es la proporción poblacional del estrato  $i$  y  $D = \frac{B^2}{Z^2}$ .

En el caso de la asignación aproximada que minimiza el costo para un valor fijo de  $v(\hat{p}_{est})$  o minimiza  $v(\hat{p}_{est})$  para un costo fijo estará dado por:

$$n_i = n \frac{N_i \sqrt{p_i q_i / c_i}}{\sum_{k=1}^L N_k \sqrt{p_k q_k / c_k}}, \text{ fórmula 3.18}$$



De igual forma que en el caso de la asignación proporcional para medias y totales, la expresión es la misma expresión que para proporciones.

$$n_i = n \left( \frac{N_i}{N} \right); i = 1, 2, \dots, L.$$

Las expresiones para determinar  $n$  son las mismas que hemos planteado para las medias y totales, con la única salvedad de que se sustituya la expresión  $\sigma_i$  por  $\sqrt{p_i q_i}$  para encontrar  $n$  en el caso de las proporciones dado que la varianza poblacional del estrato  $i$  para una proporción, es el producto de  $p_i$  por  $q_i$ , esto es:  $\sigma_i^2 = p_i q_i$

## Actividad de autoaprendizaje No. 1

- Supongamos que la calificación promedio para el examen de la clase del ejercicio que hemos resuelto anteriormente, se va a estimar de nuevo al final del año escolar. Los costos de muestreo son iguales en todos los estratos, pero las varianzas son diferentes. Encuentro la asignación óptima (Neyman) para una muestra de tamaño 50, usando los datos ya suministrados para aproximar las varianzas. Anexamos dichos datos para que sean utilizados en lo que se nos solicita:

Estrato I		Estrato II		Estrato III	
80	92	85	82	42	32
68	85	48	75	36	31
72	87	53	73	65	29
85	91	65	78	43	19
90	81	49	69	53	14
62	79	72	81	61	31
61	83	53	59	42	30
		68	52	39	32
		71	61		
		59	42		

- Utilizando los datos en la tabla del ejercicio anterior, encuentro el tamaño de muestra requerido para estimar la calificación promedio, con un límite de 4 puntos para el error de estimación. Uso la asignación proporcional.
- Repito el ejercicio anterior, ahora usando la asignación de Neyman. Comparo los resultados con la respuesta obtenida del ejercicio 2.
- Un ayuntamiento municipal está interesado en ampliar las instalaciones de un centro de atención diurna para niños con discapacidades motoras. La ampliación va a

incrementar los costos de asistencia a los niños del centro. Se va a realizar una encuesta por muestreo para estimar la proporción de familias con niños afectados que utilizarán las instalaciones ampliadas. Las familias están divididas en aquellas que usan las instalaciones y aquellas que no lo hacen. Algunas familias viven en la ciudad donde se encuentra localizado el centro, y otras viven en las áreas rurales o sus alrededores. Se utiliza muestreo aleatorio estratificado con personas en la ciudad que usan las instalaciones, personas de los alrededores que las usan, personas en la ciudad que no las usan, y personas en los alrededores que no las usan, formando los estratos 1, 2, 3, y 4, respectivamente. Aproximadamente 90% de los que usan las instalaciones y 50% de los que no las usan van a utilizar las nuevas instalaciones. Los costos por efectuar la observación de un cliente actual es de C\$4.00 y de C\$8.00 para uno que no lo es. La diferencia en el costo resulta de la dificultad para localizar a quienes no usan las instalaciones.

Registros existentes nos dan  $N_1 = 97$ ,  $N_2 = 43$ ,  $N_3 = 145$ ,  $N_4 = 68$ . encuentro el tamaño de muestra aproximado y la asignación necesaria para estimar la proporción poblacional con un límite de 0.05 para el error de estimación.

5. Se lleva a cabo la encuesta del ejercicio anterior, y se obtiene la siguiente proporción de familias que usarán las nuevas instalaciones:

$$\bar{p}_1 = 0.87,$$

$$\bar{p}_2 = 0.93,$$

$$\bar{p}_3 = 0.60,$$

$$\bar{p}_4 = 0.53$$

Con estos datos estimo la proporción poblacional  $p$ , y establezco un límite para el error de estimación. Valoro si se logró el límite deseado.

Encontraré las respuestas a esta actividad de autoaprendizaje en la página 225 al final de la unidad autoformativa III.

## **B. MUESTREO ALEATORIO POR CONGLOMERADOS**

En los tipos de muestreo analizados hasta ahora, la unidad muestral la componen los elementos de la población objeto de la investigación. Existe otro procedimiento de muestreo, también aleatorio, en el que la unidad muestral no son los individuos, sino un conjunto de individuos que, bajo determinados aspectos, se puede considerar que forman una unidad. A este tipo de muestreo se le denomina muestreo por conglomerados. Las unidades hospitalarias, los departamentos universitarios, una caja de productos terminados, constituyen ejemplos de conglomerados naturales. Existen otros tipos de conglomerados no naturales formados también por conjuntos de elementos, que pueden recibir el mismo tratamiento muestral que los casos anteriores como, por ejemplo, las urnas electorales. Hay que destacar, además, por su gran utilización, los conglomerados definidos como áreas bien delimitadas del terreno, de modo que todas las unidades últimas correspondientes al área sean las que constituyen el conglomerado. En estos casos, en vez de hablar de muestreo por conglomerados, se suele hablar de muestreo por áreas, aunque la técnica a utilizar es la misma.

### **1. Definición**

Una muestra por conglomerados es una muestra aleatoria en la cual cada unidad de muestreo es una colección, o conglomerado, de elementos.

Un ejemplo lo puede ser las manzanas en que está constituido el municipio de Estelí por ejemplo y en donde cada casa es un elemento y las 20 o más casas de cada manzana constituye lo que denominamos la unidad del conglomerado.

El muestreo por conglomerados es menos costoso que el muestreo aleatorio estratificado o irrestricto, si el costo por obtener un marco que liste todos los elementos poblacionales es muy alto o si el costo por obtener observaciones se incrementa con la distancia que separa los elementos.

Para resumir, el muestreo por conglomerados es un diseño efectivo para obtener una cantidad especificada de información al costo mínimo bajo las siguientes condiciones:

- No se encuentra disponible o es muy costoso obtener un buen marco que liste los elementos de la población, mientras que se puede lograr fácilmente un marco que liste los conglomerados.
- El costo por obtener observaciones se incrementa con la distancia que separa los elementos.

### **2. Procedimiento de selección de una muestra aleatoria por conglomerado**

**Primero:** Dividir la población objetivo en conglomerados, esto es especificar los conglomerados apropiados. Los conglomerados, deben ser tan heterogéneos (diferentes) entre ellos como sea posible, y un conglomerado debe ser muy similar a otro para poder aprovechar las ventajas económicas del muestreo por conglomerados.

**Segundo:** Una vez que los conglomerados han sido especificados se debe conformar un marco que liste todos los conglomerados de la población. Observar que en el muestreo por conglomerado, son estos los que pasan al siguiente paso y no los elementos que son los que en los otros métodos pasaban para la selección de la muestra.

**Tercero:** Entonces se selecciona una muestra irrestricta aleatoria de conglomerados de este marco mediante el uso del método aleatorio simple, aplicándosele a todos los elementos que conforman el conglomerado. La forma de operar para extraer la muestra depende, entre otras cosas, del tamaño del conglomerado. Cuando este es pequeño, se puede proceder entrevistando a todos los elementos del conglomerado. Si los conglomerados son muy grandes resulta imposible la aplicación de las entrevistas a todos y cada uno de sus elementos por lo que hay que recurrir al sub-muestreo.

La primera tarea en muestreo por conglomerados es entonces especificar los conglomerados apropiados. Los elementos dentro de un conglomerado están frecuentemente juntos físicamente, por lo que tienden a presentar características similares. Dicho de otra manera, la medición de un elemento en un conglomerado puede estar altamente correlacionada con la de otro elemento. Entonces la cantidad de información acerca de un parámetro poblacional puede no incrementarse sustancialmente al tomar nuevas mediciones dentro de un conglomerado. Ya que las mediciones cuestan dinero, un experimentador podría desperdiciar presupuesto si es que selecciona un conglomerado de gran tamaño. Sin embargo pueden ocurrir situaciones en las cuales los elementos dentro de un conglomerado son muy diferentes entre sí. En tales casos una muestra que contenga pocos conglomerados grandes puede producir una estimación muy buena de un parámetro poblacional, tal como la media.

Nótese la principal diferencia entre la construcción óptima de estratos y la construcción de los conglomerados. La diferencia es que los estratos deben ser tan homogéneos (semejantes) entre ellos, como sea posible, pero un estrato debe diferir tanto como sea posible de otro con respecto a la característica que está siendo medida. Los conglomerados, por otro lado, deben ser tan heterogéneos (diferentes) entre ellos como sea posible, y un conglomerado debe ser muy similar a otro para poder aprovechar las ventajas económicas del muestreo por conglomerados.

Una vez que los conglomerados han sido especificados se debe conformar un marco que liste todos los conglomerados de la población. Entonces se selecciona una muestra irrestricta aleatoria de conglomerados de este marco mediante el uso del método aleatorio simple ya estudiado.

El muestreo por conglomerados monoetápico, que es el que ahora interesa, se reduce en la práctica a la aplicación de los muestreos aleatorio simple, sistemático o estratificado, pero tomando como unidad de muestreo los conglomerados y no los individuos.

### ***Observemos un ejemplo tomado de la realidad.***

*La muestra que sirvió de base para este estudio se diseñó mediante muestreo aleatorio por conglomerados con las siguientes características:*

*Universo:* Las unidades clases de primero, segundo y tercer grado de los centros escolares del municipio de Madrid.

*Estratificación:* Los centros se estratificaron según su tamaño y su pertenencia al sector público o privado. Del cruce de estas dos variables se obtuvieron los 9 estratos que aparecen en el siguiente cuadro.

Tipo de Centro				
Tamaño	Público	Privado religioso	Privado no religioso	Total
Pequeño	7	154	159	320
Mediano	90	94	120	304
Grande	350	9	17	376
Total	447	257	296	1000

**Tamaño de la muestra:** Se fija en 133 unidades (clases). Estas se distribuyeron de forma proporcional al peso de cada estrato. Las 133 unidades responden a unos 4.000 alumnos.

**Selección de los elementos de la muestra:** A partir de un listado de unidades clases por centros escolares y estratos se extrajeron las unidades que deberían formar parte de la muestra, mediante muestreo aleatorio simple. Los cuestionarios se aplicaron a todos los alumnos de las clases seleccionadas.

Las muestras por conglomerados están más concentradas frente a una dispersión mucho mayor en el resto de los métodos de muestreo. Por eso los muestreos por conglomerados simplifican los procesos de extracción de la muestra, reduciendo costos y tiempo en su ejecución. En este tipo de muestreo el conocimiento del universo objeto del estudio se simplifica, dado que basta con tener un listado de conglomerados, del cual se van a seleccionar los que van a formar parte de la muestra. Son estos, los seleccionados, los que van a servir de base para recoger la información que se pretende obtener mediante el muestreo.

En resumen, se puede decir que en el muestreo por conglomerados la unidad muestral no son los individuos sino un conjunto de ellos; la muestra resultante está mucho más concentrada y se opera con un nivel de información sensiblemente inferior.

### 3. Estimaciones basada en una muestra aleatoria por conglomerado

El muestreo por conglomerados es un muestreo irrestricto aleatorio con cada unidad de muestreo conteniendo un número de elementos. Por esto los estimadores de la media poblacional  $\bar{y}$  y el total  $\tau$  son similares a los del muestreo irrestricto aleatorio. En particular la media muestral  $\bar{X}$  es un buen estimador de la media poblacional  $\bar{y}$ .

Se utilizará la siguiente notación:

$N$  = número de conglomerados en la población.

$n$  = número de conglomerados seleccionados en una muestra irrestricta aleatoria.

$m_i$  = número de elementos en el conglomerado  $i$ ,  $i = 1, \dots, N$

$\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$  = tamaño promedio del conglomerado de la muestra.

$M = \sum_{i=1}^N m_i$  = número de elementos en la población.

$\bar{M} = \frac{M}{N}$  = tamaño promedio del conglomerado en la población.

$X_i$  = total de todas las observaciones en el  $i$ -ésimo conglomerado.



#### 4. Estimador de la media poblacional $\mu$ , varianza estimada de la media y límite para el error de estimación de la media

El estimador de la media poblacional  $\bar{x}$  se calcula por:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n m_i}, \text{ fórmula 3.19}$$

La Varianza estimada de  $\bar{x}$  por medio de:  $\hat{V}(\bar{X}) = \left( \frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (X_i - \bar{X}m_i)^2}{n-1}$ , fórmula 3.20

El Límite para el error de estimación por la expresión:

$$Z\sqrt{\hat{V}(\bar{X})} = Z\sqrt{\left( \frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (X_i - \bar{X}m_i)^2}{n-1}}, \text{ fórmula 3.21}$$

Aquí  $\bar{M}$  puede ser estimado por  $\bar{m}$  si se desconoce M.

La varianza estimada en la Ecuación  $\hat{V}(\bar{X})$  para la varianza estimada de  $\bar{X}$  es sesgada y sería un buen estimador de  $V(\bar{X})$  únicamente si n fuera grande, digamos  $n \geq 20$ . El sesgo desaparece cuando los tamaños de los conglomerados  $m_1, m_2, \dots, m_N$  son iguales.

#### ***Veamos el siguiente ejercicio:***

Un fabricante de sierras de cinta quiere estimar el costo de reparación promedio mensual para las sierras que ha vendido a ciertas industrias. El fabricante no puede obtener un costo de reparación para cada sierra, pero puede obtener la cantidad total gastada en reparación y el número de sierras que tiene cada industria. Entonces decide usar muestreo por conglomerados, con cada industria como un conglomerado. El fabricante selecciona una muestra irrestricta aleatoria de  $n = 20$  de  $N = 96$  industrias a las que da servicio. Los datos sobre costo total de reparaciones por industria y el número de sierras por industria se presentan en la tabla anexa. Estime el costo promedio de reparación por sierra para el mes pasado, y establezca un límite para el error de estimación.

Industria	Número de Sierras	Costo Total de reparación para el mes pasado (en dólares)	Industria	Número de Sierras	Costo Total de reparación para el mes pasado (en dólares)
1	3	50	11	8	140
2	7	110	12	6	130
3	11	230	13	3	70
4	9	140	14	2	50
5	2	60	15	1	10
6	12	280	16	4	60
7	14	240	17	12	280
8	3	45	18	6	150
9	5	60	19	5	110
10	9	230	20	8	120

**Solución:**

$n = 20$  conglomerados que corresponden a las 20 industrias

$N = 96$  industrias que es el total de las industrias a las que la empresa les vende

$m_i$ : Número de sierras en el conglomerado  $i$ .

$x_i$ : Costo total de reparación en el conglomerado  $i$ .

Calculando el estimador de la media  $\mu$  por medio de la fórmula 3.19, tenemos:

$$\sum_{i=1}^{20} x_i = 2,565$$

$$\sum_{i=1}^{20} m_i = 130$$

$$\bar{x} = \frac{\sum_{i=1}^{20} x_i}{\sum_{i=1}^{20} m_i} = \frac{2,565}{130} = 19.73$$

19.73 es el costo promedio de reparación por sierra.

La Varianza estimada de  $\bar{x}$  es  $V(\bar{x}) = \left( \frac{N-n}{Nm\bar{m}^2} \right) \left( \frac{\sum_{i=1}^n (x_i - \bar{x}m_i)^2}{n-1} \right)$ , dado por la fórmula 3.20 y que  $M$  es

desconocida,  $\bar{M}$ , debe ser estimada por  $\bar{m}$

$$\bar{m} = \frac{1}{20} \sum_{i=1}^{20} m_i = \frac{1}{20} \times 130 = 6.5.$$

Realizando los cálculos:

$$\begin{aligned} \sum_{i=1}^{20} (x_i - \bar{x}m_i)^2 &= \sum_{i=1}^{20} (x_i^2 - 2x_i\bar{x}m_i + \bar{x}^2 m_i^2) = \sum_{i=1}^{20} x_i^2 - 2\bar{x} \sum_{i=1}^{20} x_i m_i + \bar{x}^2 \sum_{i=1}^{20} m_i^2 = \\ &= 460,225 - 2 \times 19.73 \times 22,285 + (19.73)^2 \times 1,118 = 16,066 \end{aligned}$$

Donde:

$$\sum_{i=1}^{20} x_i^2 = 460,225$$

$$\sum_{i=1}^{20} m_i^2 = 1,118$$

$$\sum_{i=1}^{20} x_i m_i = 22,285$$

Por lo que:  $V(\bar{x}) = \left( \frac{96-20}{96 \times 20 \times 6.5^2} \right) \times \frac{16,066}{20-1} = 0.7922$

El límite para el error de estimación con 95% de nivel de confianza se calcula con la fórmula 3.21 y será:

$$z\sqrt{V(\bar{x})} = 1.96\sqrt{0.7922} = 1.74$$

Por lo que  $\mu \rightarrow \bar{x} \pm z\sqrt{V(\bar{x})} = 19.73 \pm 1.74$ ; entonces  $\mu \rightarrow (17.99, 21.47)$

De lo anterior deducimos que el costo promedio de reparación por sierra para el mes pasado, se estima entre 17.99 y 21.47 unidades monetarias a un nivel de confianza del 95%.

## 5. Estimador del total poblacional $\tau$ , varianza estimada del total y límite para el error de estimación del total poblacional

El total poblacional  $\tau$  es ahora  $M\bar{X}$  porque  $M$  denota el número total de elementos en la población. Por ende, como en el muestreo irrestricto aleatorio,  $M\bar{X}$  proporciona un estimador  $\tau$ .

El estimador del total poblacional  $\tau$  estará dado por:  $M\bar{X} = M \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n m_i}$ , fórmula 3.22

La Varianza estimada de  $M\bar{X}$  será:  $\hat{V}(M\bar{X}) = M^2 \hat{V}(\bar{X}) = N^2 \left( \frac{N-n}{Nn} \right) \frac{\sum_{i=1}^n (X_i - \bar{X}m_i)^2}{n-1}$ , fórmula 3.23

El Límite para el error de estimación, será:  $Z\sqrt{\hat{V}(M\bar{X})} = Z\sqrt{N^2 \left( \frac{N-n}{Nn} \right) \frac{\sum_{i=1}^n (X_i - \bar{X}m_i)^2}{n-1}}$ , fórmula 3.24

Nótese que el estimador  $M\bar{X}$  es útil únicamente si se conoce el número de elementos  $M$  en la población.

Frecuentemente el número de elementos en la población no es conocido en problemas donde el muestreo por conglomerados es apropiado. Entonces no podemos usar el estimador  $M\bar{X}$ , pero podemos formar otro estimador del total poblacional que no depende de  $M$ . La cantidad  $\bar{X}_t$ , dada por:

$$\bar{X}_t = \frac{1}{n} \sum_{i=1}^n X_i, \text{ fórmula 3.25}$$

es el promedio de los totales de conglomerados para los  $n$  conglomerados muestreados. Es por esto que  $\bar{X}_t$  es un estimador insesgado del promedio de los  $N$  totales de conglomerados en la población. Por el mismo razonamiento empleado,  $N\bar{X}_t$  es un estimador insesgado de la suma de los totales de conglomerados o, equivalentemente, del total poblacional  $\tau$ .

Por ejemplo es altamente improbable que se conozca el número de adultos varones en una ciudad, por lo que el estimador  $N\bar{X}_t$  tendrá que ser usado en lugar de  $M\bar{X}$  para estimar  $\tau$ .

El estimador del total poblacional  $\tau$ , el cual no depende de  $M$ , se encuentra por medio de:



$$N\bar{X}_t = \frac{N}{n} \sum_{i=1}^n X_i, \text{ fórmula 3.26}$$

La Varianza estimada de  $N\bar{x}$ , será:  $V(N\bar{X}_t) = N^2 V(\bar{X}_t) = N^2 \left( \frac{N-n}{Nn} \right) \frac{\sum_{i=1}^n (X_i - \bar{X}_t)^2}{n-1}$ , fórmula 3.27

El Límite para el error de estimación, estará dado por:

$$Z\sqrt{V(N\bar{X}_t)} = Z\sqrt{N^2 \left( \frac{N-n}{Nn} \right) \frac{\sum_{i=1}^n (X_i - \bar{X}_t)^2}{n-1}}, \text{ fórmula 3.28}$$

Si existe una gran cantidad de variación entre los tamaños de los conglomerados y si los tamaños están altamente correlacionados con los totales de conglomerados, la varianza de  $N\bar{X}_t$  es generalmente mayor que la varianza de  $M\bar{X}$ . El estimador  $N\bar{X}_t$  no usa la información proporcionada por los tamaños de los conglomerados  $m_1, m_2, \dots, m_n$  y por esto puede ser menos preciso.

### ***Veamos los siguientes ejercicios:***

*Para los datos del ejercicio de las sierras que hemos resuelto anteriormente, estime la cantidad total gastada por las 96 industrias en la reparación de sierras. Establezca un límite para el error de estimación.*

### **Solución:**

*Para encontrar la cantidad total gastada por las 96 industrias, se puede calcular por medio de  $M\bar{x}$ , pero no conocemos el valor de M: número de elementos en la población, por tal razón utilizamos otro estimador que es  $\bar{x}_t = \frac{1}{n} \sum_{i=1}^n x_i$ , que es el promedio de los totales de conglomerados para los  $n$  conglomerados muestreados.*

*Aplicando las fórmulas 3.26 y 3.27, el estimador del total poblacional  $\bar{T}$ , el cual no depende de M es entonces  $N\bar{x}_t = \frac{N}{n} \sum_{i=1}^n x_i = \frac{96}{20} \sum_{i=1}^{20} x_i = \frac{96}{20} (2,565) = 12,312$ , y su varianza estimada es:*

$$V(N\bar{x}_t) = N^2 \left( \frac{N-n}{Nn} \right) \left( \frac{\sum_{i=1}^n (x_i - \bar{x}_t)^2}{n-1} \right) = (96)^2 \times \left( \frac{96-20}{96 \times 20} \right) \left( \frac{131,263.75}{19} \right) = 2,520,264$$

*Dado que:  $\sum_{i=1}^n (x_i - \bar{x}_t)^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2 = 460,225 - \frac{1}{20} \times (2,565)^2 = 131,263.75$*

*El límite para el error de estimación con un nivel de confianza del 95%, se encuentra aplicando la fórmula 3.28, resultándonos:  $z\sqrt{V(N\bar{x}_t)} = 1.96\sqrt{2,520,264} = 3,111.57$*

*Por lo que:*

$$\tau \rightarrow N\bar{x}_t \pm z\sqrt{V(N\bar{x}_t)} = 12,312 \pm 3,111.57$$

$$\tau \rightarrow (9,200.43 ; 15,423.57)$$

La cantidad total gastada por las 96 industrias en la reparación de sierras, se estima entre 9,200.43 y 15,423.57 unidades monetarias a un nivel de confianza del 95%.

### **Veamos otro ejercicio:**

Después de verificar sus registros de ventas, el fabricante de sierras del ejercicio anterior, se percató de que ha vendido un total de 710 sierras a esas industrias. Usando esta información adicional, estime la cantidad total gastada en reparación de sierras por estas industrias, y establezca un límite para el error de estimación.

### **Solución:**

Observemos que ahora nos están dando la información de  $M$ , con  $M = 710$ , y  $\bar{x} = 19.73$ , que se ha calculado en el primer ejercicio resuelto anteriormente, se tiene que:  $\tau = M\bar{x} = 710 \times 19.73 = 14,008.30$  aplicando la fórmula 3.22

La varianza estimada de  $M\bar{x}$  aplicando la fórmula 3.23 es:

$$V(M\bar{x}) = N^2 \left( \frac{N-n}{Nn} \right) \left( \frac{\sum (x_i - \bar{x}m_i)^2}{n-1} \right) = (96)^2 \times \left( \frac{96-20}{96 \times 20} \right) \times \left( \frac{16,066}{20-1} \right) = 308,467.20$$

El límite para el error de estimación con 95% de nivel de confianza será:  $z\sqrt{V(M\bar{x})} = 1.96 \times \sqrt{308,467.20} = 1,088.58$ , al aplicar la fórmula 3.24, por lo que se tiene que el total:

$$\tau \rightarrow \bar{\tau} \pm z\sqrt{V(M\bar{x})} = 14,008.30 \pm 1,088.58$$

$$\tau \rightarrow (12,919.72; 15,096.88)$$

Con base a los nuevos datos suministrados por el ejercicio, el total gastado se estima entre 12,919.72 y 15,096.88 unidades monetarias para un nivel de confianza del 95%. Observar que en este ejercicio se nos ha brindado la información de las 710 sierras vendidas a las industrias.

¡Compare los resultados con los del ejercicio anterior!. ¿Qué opinión le merece?

## **6. Cálculo del tamaño de la muestra para la media poblacional y el total poblacional**

El tamaño de muestra aproximado requerido para estimar  $\bar{\mu}$  con un límite  $B$  para el error de estimación se calcula por medio de:

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2}, \text{ fórmula 3.29}$$

Donde  $\sigma_c^2$  es estimado por  $s_c^2$  (el subíndice  $c$  es indicativo de conglomerado) y  $D = \frac{B^2 \bar{M}}{Z^2}$

**Veamos el siguiente ejercicio:**

El mismo fabricante de sierras en base a los datos suministrados por el primer ejercicio resuelto, quiere estimar el costo de reparación promedio por sierra para el mes siguiente. ¿Cuántos conglomerados debe seleccionar en la muestra si quiere que el límite para el error de estimación sea menor que C\$2.00?

**Solución:**

La expresión que relaciona el tamaño de los conglomerados, está dada por la fórmula 3.29:

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2}, \text{ Donde } \sigma_c^2 \text{ es estimado por } s_c^2 \text{ y } D = \frac{B^2 \bar{M}^2}{Z^2}$$

$$B = 2$$

$$\bar{M} = \bar{m} = 6.5$$

$$Z = 1.96$$

$$s_c^2 = \frac{\sum (x_i - \bar{m}_i)^2}{n-1} = \frac{16,066}{19} = 845.5789$$

$$D = \frac{(2)^2 (6.5)^2}{(1.96)^2} = 43.99$$

$$n = \frac{96(845.58)}{96(43.99) + 845.58} = 16.01$$

$$n = 16$$

El total de conglomerados que se debe seleccionar para estimar el costo de reparación promedio por sierra con un error de estimación menor de C\$2.00 para el mes, es de  $n = 16$  a un nivel de confianza del 95%.

El tamaño de muestra aproximado requerido para estimar  $\tau$ , usando  $M\bar{X}$  con un límite B para el error de estimación, estará dado por:

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2}, \text{ fórmula 3.30}$$

Donde  $\sigma_c^2$  es estimado por  $s_c^2$  y  $D = \frac{B^2}{Z^2 N^2}$

El tamaño de muestra aproximado requerido para estimar  $\tau$ , usando  $N\bar{X}_t$  con un límite B para el error de estimación estará dado por:

$$n = \frac{N\sigma_t^2}{ND + \sigma_t^2}, \text{ fórmula 3.31}$$

Donde  $\sigma_t^2$  es estimado por  $s_t^2$  y  $D = \frac{B^2}{Z^2 N^2}$

## 7. Estimador de una proporción poblacional, varianza estimada de la proporción y límite para el error de estimación de la proporción poblacional.

Supóngase que un experimentador desea estimar una proporción poblacional, o fracción, tal como la proporción de casas en un municipio con inadecuado servicio de cañería de agua potable, o la proporción de presidentes de empresas que son universitarios graduados. El mejor estimador de la proporción poblacional  $p$  es la proporción muestral  $\hat{p}$ . Sea  $a_i$ , el número total de elementos en el conglomerado  $i$  que poseen la característica de interés. Entonces, la proporción de elementos en la muestra de  $n$  conglomerados que poseen la característica de interés es dada por:

$$\hat{p} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i}$$

donde  $m_i$  es el número de elementos en el  $i$ -ésimo conglomerado,  $i = 1, 2, \dots, n$ . Nótese que  $\hat{p}$  tiene la misma forma de  $\bar{X}$ , excepto que  $X_i$  es reemplazado por  $a_i$ . La varianza estimada de  $\hat{p}$  es similar a la de  $\bar{X}$ .

El estimador de la proporción poblacional  $p$  será entonces:  $\hat{p} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i}$ , fórmula 3.32

La Varianza estimada de  $\hat{p}$  estará dada por:  $\hat{V}(\hat{p}) = \left( \frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (a_i - \hat{p}m_i)^2}{n-1}$ , fórmula 3.33

El Límite para el error de estimación por:  $Z\sqrt{\hat{V}(\hat{p})} = Z\sqrt{\left( \frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (a_i - \hat{p}m_i)^2}{n-1}}$ , fórmula 3.34

La fórmula de varianza es un buen estimador únicamente cuando la muestra de tamaño  $n$  es grande, digamos  $n \geq 20$ . Si  $m_1 = m_2 = \dots = m_N$ , entonces  $\hat{p}$  es un estimador insesgado de  $p$ , y la  $\hat{V}(\hat{p})$ , que se muestra en la ecuación de la varianza estimada de  $\hat{p}$ ; es un estimador insesgado de la varianza real de  $\hat{p}$  para cualquier tamaño de muestra.

### Veamos el siguiente ejercicio:

Una industria está considerando la revisión de su política de jubilación y quiere estimar la proporción de empleados que apoyan la nueva política. La industria consiste de 87 plantas separadas localizadas en todo Estados Unidos. Ya que los resultados deben ser obtenidos rápidamente y con poco dinero, la industria decide usar muestreo por conglomerados, con cada planta como un conglomerado. Se selecciona una muestra irrestricta aleatoria de 15 plantas y se obtienen las opiniones de los empleados en estas plantas a través de un cuestionario. Los resultados se presentan en la tabla anexa. Estime la proporción de empleados en la industria que apoyan la nueva política de jubilación y establezca un límite para el error de estimación.

Planta	Número de empleados	Número de empleados que apoyan la nueva política	Planta	Número de empleados	Número de empleados que apoyan la nueva política
1	51	42	9	73	54
2	62	53	10	61	45
3	49	40	11	58	51
4	73	45	12	52	29
5	101	63	13	65	46
6	48	31	14	49	37
7	65	38	15	55	42
8	49	30			

### Solución:

Los datos que tenemos son:  $N = 87$   $n = 15$

Los cálculos a realizar aplicando la fórmula 3.32 serán para determinar la proporción poblacional como a continuación se realizan:

$$\sum_{i=1}^{15} a_i = 646$$

$$\sum_{i=1}^{15} m_i = 911$$

$$\hat{p} = \frac{\sum a_i}{\sum m_i} = \frac{646}{911} = 0.7091$$

La Varianza estimada de  $\hat{p}$  se calcula aplicando la fórmula 3.33 
$$v(\hat{p}) = \left( \frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum (a_i - \hat{p}m_i)^2}{n-1}$$

como  $\bar{M}$  no es conocida, se sustituye por  $\bar{m} = \frac{1}{n} \sum m_i = \frac{911}{15} = 60.733...$

$$\sum (a_i - \hat{p}m_i)^2 = \sum a_i^2 - 2\hat{p} \sum a_i m_i + \hat{p}^2 \sum m_i^2$$

$$\sum a_i^2 = 29,104$$

$$\sum a_i m_i = 40730$$

$$\sum m_i^2 = 58,075$$

$$\sum (a_i - \hat{p}m_i)^2 = 29,104 - 2(0.7091)(40730) + (0.7091)^2(58,075) = 542.15$$

$$v(\hat{p}) = \left( \frac{87-15}{(87)(15)(60.7333)^2} \right) \frac{542.05}{14} = 0.00057924$$

El límite para el error de estimación se calcula aplicando la fórmula 3.34:

$$Z\sqrt{v(\hat{p})} = 1.96\sqrt{0.00057924} = 0.0472$$

$$\hat{p} \rightarrow \hat{p} \pm Z\sqrt{v(\hat{p})} = 0.7091 \pm 0.0472$$

$$\hat{p} \rightarrow (0.6619, \quad 0.7563)$$

La proporción de empleados en la industria que apoyan la nueva política de jubilación se estima entre el 66.2% y el 75.6% para un nivel de confianza del 95%.

## 8. Cálculo del tamaño de la muestra para las proporciones:

La estimación de la proporción poblacional  $p$ , con un límite de  $B$  unidades para el error de estimación, implica que el experimentador quiere que:

$$Z\sqrt{V(\hat{p})} = B$$

Esta ecuación puede ser resuelta para  $n$ , y la solución es similar a la ecuación del tamaño de la muestra requerido para estimar  $\mu$  con un límite  $B$ .

Esto es:  $n = \frac{N\sigma_c^2}{ND + \sigma_c^2}$ , fórmula 3.35

donde  $D = B^2 \bar{M}^2 / Z^2$ , y  $\sigma_c^2$  se estima por:  $S_c^2 = \frac{\sum_{i=1}^n (a_i - \hat{p} m_i)^2}{n-1}$

### **Veamos el siguiente ejercicio:**

La industria del ejercicio resuelto anteriormente, modificó su política de jubilación después de obtener los resultados de la encuesta. Ahora se quiere estimar la proporción de empleados a favor de la política modificada. ¿Cuántas plantas deben ser muestreadas para tener un límite de 0.08 para el error de estimación? Use los datos del ejercicio anterior para aproximar los resultados de la nueva encuesta.

### **Solución:**

Estamos interesados en encontrar el número de conglomerados para realizar las estimaciones con un límite para el error de estimación de 0.08 en la proporción que se estima de la política de jubilación modificada, por lo que aplicamos la fórmula 3.35 y tenemos:

$$n = \frac{N\sigma_c^2}{ND + \sigma_c^2} \text{ donde } D = \frac{B^2 \bar{m}^2}{Z^2} = \frac{(0.08)^2 (60.733)^2}{(1.96)^2} = 6.145$$

$$\text{y } \sigma_c^2 \text{ se estima con } S_c^2 = \frac{542.15}{14} = 38.725$$

$$n = \frac{87(38.725)}{(87)(6.145) + 38.725} = 5.87$$

$$n = 6 \text{ conglomerados}$$

Por los resultados obtenidos, deben ser muestreadas seis plantas ( $n = 6$ ) para realizar una estimación con un límite para el error de estimación de 0.08 con un nivel de confianza del 95%.

## Actividad de autoaprendizaje No.2

- Se diseña una encuesta económica para estimar la cantidad promedio gastada en servicios para el hogar en una ciudad. Ya que no se encuentra disponible una lista de hogares, se usa muestreo por conglomerados, con divisiones (barrios) formando los conglomerados. Se selecciona una muestra aleatoria de 20 barrios de la ciudad de un total de 60. Los entrevistadores obtienen el costo de los servicios de cada hogar dentro de los barrios seleccionados; los costos totales se muestran en la tabla anexa. Estime la cantidad promedio de gastos en servicios por hogar en la ciudad y establezca un límite para el error de estimación.

Barrio Muestreado	Número de hogares	Cantidad total gastada en servicios	Barrio Muestreado	Número de hogares	Cantidad total gastada en servicios
1	55	C\$2210	11	73	C\$2930
2	60	2390	12	64	2470
3	63	2430	13	69	2830
4	58	2380	14	58	2370
5	71	2760	15	63	2390
6	78	3110	16	75	2870
7	69	2780	17	78	3210
8	58	2370	18	51	2430
9	52	1990	19	67	2730
10	71	2810	20	70	2880

- En la encuesta del Ejercicio 1 se desconoce el número de hogares en la ciudad. Estime la cantidad total gastada en servicios por todos los hogares de la ciudad y establezca un límite para el error de estimación.
- La encuesta económica del Ejercicio 1 se va a llevar a cabo en una ciudad vecina de estructura similar. El objetivo es estimar la cantidad total gastada en servicios por los hogares de la ciudad, con un límite de \$ 5000 para el error de estimación. Use los datos del ejercicio 1 para encontrar el número aproximado de conglomerados que se necesitan para obtener este límite.
- Un inspector quiere estimar el peso promedio de llenado para cajas de cereal empacadas en una fábrica. El cereal está en paquetes que contienen doce cajas cada uno. El inspector selecciona aleatoriamente cinco y mide el peso de llenado de cada caja en los paquetes muestreados, con los resultados (en onzas) que se muestran en la tabla acompañante. Estime el peso promedio de llenado para las cajas empacadas por esta fábrica, y establezca un límite para el error de estimación. Suponga que el número total de cajas empacadas por la fábrica es lo suficientemente grande para que no se tome en cuenta la corrección por población finita (Recordar que no se considera el factor de corrección por población finita).

Paquete	Onzas de llenado											
1	16.1	15.9	16.1	16.2	15.9	15.8	16.1	16.2	16.0	15.9	15.8	16.0
2	15.9	16.2	15.8	16.0	16.3	16.1	15.8	15.9	16.0	16.1	16.1	15.9
3	16.2	16.0	15.7	16.3	15.8	16.0	15.9	16.0	16.1	16.0	15.9	16.1
4	15.9	16.1	16.2	16.1	16.1	16.3	15.9	16.1	15.9	15.9	16.0	16.0
5	16.0	15.8	16.3	15.7	16.1	15.9	16.0	16.1	15.8	16.0	16.1	15.9

Encontraré las respuestas a esta actividad de autoaprendizaje en la página 225 al final de la unidad autoformativa III.

## Resumen final de la Unidad Autoformativa III

En esta unidad autoformativa del módulo de Técnicas de Recopilación de la Información, hemos estudiado otros dos métodos de muestreos probabilísticos, el aleatorio estratificado y el aleatorio por conglomerado.

En cada método de muestreo, hemos definido y establecido el procedimiento de selección de la muestra, procedimos a establecer las estimaciones de la media, el total y la proporción poblacional, así como el cálculo de las varianzas y los límites de error de estimación y cálculo según el tamaño de muestras.

En el muestreo estratificado, conocimos que se parte de una población objeto de estudio y se ordenan los elementos de la población en grupos disjuntos que conforman los estratos, para lo cual se establece claramente el criterio de estratificación. En cada estrato se persigue que las unidades que lo conforman sean lo más homogéneas entre sí y que sean heterogéneas al compararlas con los otros estratos, esto permite reducir las varianzas de las estimaciones, traduciéndose en ganancias en precisión y en costos.

En el muestreo por conglomerados comprendimos que ya no son las unidades o elementos individuales los que pasan a establecer la muestra como en los otros tres métodos, sino que cada unidad por muestrearse es un conjunto de elementos.

Vimos que el muestreo por conglomerados es efectivo cuando no se tiene la posibilidad de construir un marco muestral de los elementos individuales, pero sí de los conglomerados (o conjunto de elementos). La condición que debe estar presente en los conglomerados, es que en ellos deben ser tan diferentes los elementos que le conforman y deben ser muy similares a los otros conglomerados, situación contraria a la conformación de los estratos.

De esta forma, hemos concluido lo propuesto en el programa de la asignatura siendo el aprendizaje de su componente, importante en su formación profesional, que le apoyará en el estudio de los módulos de "Técnicas de Investigación" e "Investigación de Mercados" que estudiará en el futuro inmediato.



## **Evaluación Final de la Unidad autoformativa III**

La base de datos que se anexa, refleja las notas finales así como el sexo de una población de  $N=149$  estudiantes que han cursado la asignatura X en un cuatrimestre en la Universidad Centroamericana.

Se le solicita:

1. Realizar dos estratos conforme el criterio sexo.
2. Si estamos interesados en una muestra de tamaño  $n = 30$  y  $\sigma_1^2 = 281.03$   $\sigma_2^2 = 378.65$  que valores tendrán  $n_1$ :Tamaño de muestra para mujeres y  $n_2$ :Tamaño de muestra para varones, conforme a la asignación de Neyman.
3. Seleccionar las muestras para los tamaños de estratos  $n_1$  y  $n_2$  encontrados en el punto anterior. Recordar que el procedimiento que se aplica sobre cada población es el aleatorio simple.
4. Encontrar la media muestral (nota promedio) por cada estrato.
5. Estimar la media poblacional, la varianza de la media y el límite para el error de estimación de la media con un nivel de confianza del 95%.
6. Establecer el intervalo de confianza para la media poblacional.
7. Comparar el valor real obtenido de la media poblacional con el intervalo de confianza encontrado en el punto anterior. ¿Qué puede concluir?.
8. Calcule el tamaño de la muestra para la muestra poblacional con un nivel de confianza del 95% y con un error máximo permisible de 5 puntos. ¿ De que tamaño serán las muestras de los estratos 1 y 2 conforme la asignación de Neyman?.

Base de datos población Nota Final de la asignatura X ordenados por sexo

Caso	Sexo	Nota Final	Caso	Sexo	Nota Final	Caso	Sexo	Nota Final
1	F	46	58	F	79	108	M	63
2	F	49	59	F	72	109	M	78
3	F	89	60	F	78	110	M	85
4	F	96	61	F	80	111	M	62
5	F	26	62	F	76	112	M	62
6	F	60	63	F	94	113	M	59
7	F	87	64	F	94	114	M	33
8	F	80	65	F	84	115	M	91
9	F	83	66	F	77	116	M	75
10	F	84	67	F	81	117	M	86
11	F	92	68	F	26	118	M	72
12	F	81	69	F	73	119	M	79
13	F	71	70	F	90	120	M	75
14	F	77	71	F	74	121	M	81
15	F	33	72	F	88	122	M	89
16	F	85	73	F	70	123	M	30
17	F	70	74	F	73	124	M	98
18	F	65	75	F	71	125	M	77
19	F	51	76	F	92	126	M	75
20	F	91	77	F	71	127	M	73
21	F	95	78	F	77	128	M	73
22	F	64	79	F	75	129	M	13
23	F	98	80	F	62	130	M	92
24	F	20	81	F	70	131	M	86
25	F	58	82	F	71	132	M	65
26	F	73	83	F	70	133	M	80
27	F	66	84	F	76	134	M	90
28	F	90	85	F	24	135	M	72
29	F	91	86	F	67	136	M	60
30	F	73	87	F	42	137	M	80
31	F	82	88	F	58	138	M	38
32	F	91	89	F	65	139	M	13
33	F	83	90	F	70	140	M	60
34	F	81	91	F	71	141	M	74
35	F	84	92	F	65	142	M	63
36	F	83	93	F	95	143	M	60
37	F	71	94	F	96	144	M	77
38	F	73	95	F	94	145	M	48
39	F	73	96	F	61	146	M	65
40	F	92	97	F	61	147	M	60
41	F	73	98	F	85	148	M	62
42	F	71	99	F	47	149	M	88
43	F	62	100	F	95			
44	F	78	101	F	71			
45	F	84	102	F	95			
46	F	98	103	F	90			
47	F	96	104	F	85			
48	F	83	105	F	85			
49	F	76	106	F	86			
50	F	96	107	F	60			
51	F	81						
52	F	86						
53	F	82						
54	F	89						
55	F	53						
56	F	54						
57	F	92						

Media Varones	68.1429
Var. Poblac. Varones	378.6463
N2:	42

Media Damas	74.9813
Var. Poblac. Damas	281.0277
N1:	107

## Hoja de respuestas

### Respuestas a la actividad de autoaprendizaje No. 1

$$n = 50$$

$$\begin{aligned} 1. \quad n_1 &= 12 \\ n_2 &= 20 \\ n_3 &= 18 \end{aligned}$$

$$2. \quad n = 31$$

$$3. \quad n = 31$$

$$n = 154$$

$$n_1 = 30$$

$$\begin{aligned} 4. \quad n_2 &= 14 \\ n_3 &= 75 \\ n_4 &= 35 \end{aligned}$$

$$\begin{aligned} 5. \quad \bar{p} &= 0.701 \\ z\sqrt{V(\bar{p})} &= 0.051 \end{aligned}$$

### Respuestas a la actividad de autoaprendizaje No. 2

$$\begin{aligned} 1. \quad x &= \frac{\sum x_i}{\sum m_i} = \frac{52,340}{1,303} = 40.17 \\ z\sqrt{V(x)} &= 1.96\sqrt{0.1025461} = 0.63 \end{aligned}$$

$$\begin{aligned} 2. \quad Nx_i &= 60 \times 2,617 = 157,020 \\ z\sqrt{V(x)} &= 1.96\sqrt{11,998,863.16} = 6,789.32 \end{aligned}$$

$$3. \quad n = 29$$

$$\begin{aligned} 4. \quad \bar{x} &= 16.005 \\ z\sqrt{V(\bar{x})} &= 0.0729 \end{aligned}$$

## Glosario de términos

**Afijación:** Acto de establecer el número de elementos pertenecientes a un estrato que deben ser incluidos en la muestra.

**Afijación óptima:** Afijación proporcional que es ponderada por la varianza de la variable en estudio dentro del estrato.

**Afijación proporcional:** Afijación en la que el número de elementos es proporcional al tamaño del estrato.

**Afijación simple:** Afijación en la que el número de elementos es igual para cada estrato.

**Conglomerado:** Conjunto o grupo de elementos disjuntos en una población y que constituyen una unidad muestral.

**Estrato:** Subconjunto de la población con características propias que deseamos representar en la muestra.

**Muestreo aleatorio estratificado:** Técnica que asegura que los estratos de la población mantienen una determinada representación en la muestra.

**Muestreo aleatorio por conglomerados:** Técnica en la que lo que se muestrea son conjuntos de elementos —conglomerados— en vez de los elementos mismos. Se utiliza cuando las poblaciones son muy numerosas.

## **Bibliografía**

Mason D. B. y Lind D. A. Estadística para administración y economía. 2000. Editorial Alfaomega, octava edición, México.

Berenson M. L. y Levine D.M. Estadística básica en Administración, Conceptos y aplicaciones. 1992. Editorial PHH, cuarta edición, México.

Sheaffer R., Mendenhal W., y Ott L. Elementos de muestreo. 1987. Editorial Iberoamericana, tercera edición, México.

Koroliuk V. S. Manual de la teoría de probabilidades y estadística matemática. 1981. Editorial Mir, primera edición, URSS.

Zajarov V. K., Sevastianov B. A. Y Chistiakov V. P. Teoría de las probabilidades. 1983. Editorial Mir, primera edición, URSS.

Cohran W. G. Técnicas de muestreo. 1995. Editorial CECSA, primera edición, México.

Mendenhall W. Y Reinmuth J. Estadística para administración y economía. 1995. Editorial Iberoamericana, tercera edición, México.

# va más lejos



Estimado/a estudiante :

La Universidad Centroamericana se complace en tenerlo/a como estudiante de la modalidad de educación a distancia e integrante de la comunidad educativa UCA. Al configurar este nuevo entorno de enseñanza y aprendizaje le hemos preparado específicamente este texto o módulo autoformativo, con un enfoque eminentemente pedagógico, para que usted por sí mismo/a y de manera responsable e independiente, logre con facilidad los aprendizajes de la unidad curricular que desarrolla.

El texto que ahora tiene en sus manos le proporciona información relevante para su proceso de aprendizaje. En él encontrará holísticamente integrados los componentes del proceso didáctico: objetivos, contenidos, actividades, sistema de autoevaluación y procesos de retroalimentación para la autorregulación de sus aprendizajes.

Apoyarán el autoaprendizaje del módulo, profesores tutores y compañeros de grupo que periódicamente se reunirán con usted en la universidad, para aclarar, profundizar, perfeccionar y validar los conocimientos teóricos, prácticos y actitudinales obtenidos en la solución de la propuesta educativa de este módulo. De igual manera contará con asesorías individuales para reorientar y reafirmar los resultados de su estudio.

Esperamos que los logros sean los que usted, la UCA y la sociedad nicaragüense esperan. Reciba nuestro saludo y nuestra bienvenida

Rosa Amelia Ruiz  
Directora de Educación a Distancia



Universidad Centroamericana  
Dirección de Educación a Distancia  
Msc. Rosa Amelia Ruiz  
PBX: 278 3923 - 27 ext. 309  
E-mail: amelia@ns.uca.edu.ni  
www.uca.edu.ni

Facultad de Ciencias Económicas y Empresariales  
Coordinación de Educación a Distancia  
Msc. Sandra Palacios Rodríguez  
PBX: 278 3923 - 27 ext. 130  
E-mail: spr@ns.uca.edu.ni

Facultad de Ciencias Jurídicas  
Coordinación de Educación a Distancia  
Msc. Cristian Alberto Robleto Arana  
Teléfono: 278 6509  
PBX: 278 3923 - 27 ext. 319  
E-mail: robleto@ns.uca.edu.ni

